Boost 4.0
BIG DATA FOR FACTORIES

Big Data Value Spaces for Competitiveness of European Connected Smart Factories 4.0

Horizon 2020 EU Grant Agreement 780732

Title	D2.3 – Pilots description, adaptations and executive plans v1
Document Owners	INNO
Contributors	ALL PARNTERS
Dissemination	Public
Date	15/1/2019
Version	V1.3

Version history

12/11/2018	Table of Contents
29/11/2018	1st content proposal
11/12/2018	Updated with INTRA review
12/12/2018	Updated with CARSA review
12/12/2018	Updated with VW review
02/01/2019	Contributions from FIAT
11/01/2019	Reviewed by Whirlpool
15/01/2019	Final version

Document Fiche

Authors	Oscar Lazaro – INNO, Carmen Polcaro – INNO, Jesús Alonso -INNO
Internal Reviewers	INTRA, CARSA
Workpackage	WP2
Task	T2.3
Nature	Report
Dissemination	Public

Project Partners

Participant organisation name	Acronym
Asociación de Empresas Tecnológicas Innovalia	INNO
Volkswagen Autoeuropa, Lda *	VWAE
Visual Components	VIS
Automatismos y Sistemas de Transporte Interno S.A.U.	ASTI
Telefónica Investigación y Desarrollo SA	TID
Volkswagen AG. *	VW
UNINOVA	UNINO
FILL GmbH. *	FILL
TTTECH Computertechnik AG	TTT
RISC Software GmbH	RISC
PHILIPS Consumer Lifestyle B.V. *	PCL
PHILIPS Electronics Nederland	PEN
Interuniversitair Micro-Electronicacentrum VZW	IMEC
Centro Ricerche Fiat S.C.p.A. *	CRF
SIEMENS S.p.A.	SIEMENS
Prima Industries S.p.A	PRIMA
Politecnico di Milano	POLIMI
AUTOTECH ENGINEERING, AIE *	GESTAMP
Fundació Privada I2CAT, Internet I Innovació Digital A Catalunyai2cat	I2CAT
TRIMEK S.A.	TRIMEK
CAPVIDIA N.V,	CAPVIDIA
Volvo Lastvagnar AB *	VOLVO
Chalmers Tekniska Hoegskola AB	CHAL
Whirlpool EMEA SpA *	WHIR
SAS Institute Srl	SAS
Benteler Automotive GmbH *	BAT
It.s OWL Clustermanagement	OWL
Fraunhofer Gesellschaft Zur Foerderung Der Angewandten Forschung E.V.	FhG
Atlantis Engineering	AE
Agie Charmilles New Technologies SA *	+GF+
Ecole Polytechnique Federale De Lausanne	EPFL
Institut Für Angewandte Systemtechnik Bremen GmbH	ATB
Rheinische Friedrich-Wilhelms-Universitat Bonn	UBO
Ethniko Kentro Erevnas Kai Technologikis Anaptyxis (CERTH)	CERTH

The University of Edinburgh	UED
Institute Mines Telecom	IMT
International Data Spaces e.V.	IDSA
FIWARE Foundation e.V	FF
GEIE ERCIM EEIG	ERCIM
IBM ISRAEL – Science and Technology LTD	IBM
ESI Group	ESI
Eneo Tecnología, S.L	ENEO
Software Quality Systems S.A.	SQS
Consultores de Automatización y Robótica S.A.	CARSA
INTRASOFT International	INTRA
United Technologies Research Centre Ireland, Ltd *	UTRC-I
Fratelli Piacenza S.p.A. *	PIA
RiaStone - Vista Alegre Atlantis SA *	RIA
Unparallel Innovation, Lda	UNP
Gottfried Wilhelm Leibniz Universität Hannover	LUH
Lighthouse Factory 4.0	LHF 4.0

*LHF 4.0 – Lighthouse Factory 4.0 * RF – Replication Factory 4.0

Executive Summary

The present document is the first report of the activities developed in task T2.3. This task aims to coordinate all the pilots and present, in this deliverable, each of the implantations planned in the Boost 4.0 Project.

First two trials, Volkswagen and FILL, focus on the Smart Digital Engineering Big Data. In particular, Volkswagen will introduce big data analysis in the process to design the moulds for the casing process, while FILL will work to improve their engineering processes with machine data obtained during the operation phase into the design of new machines.

The second area Smart Production Planning and Management is the main focus point for Vokswagen Autoeuropa and GF pilots. In the first one the main goal is to fully integrate the material flow, from the reception to the point of fit. In the second one, the ambition is to exploit the opportunities provided by Boost 4.0 enablers for a zero-defect factory for milling spindles for GF machines.

Smart Operations and Digital Workspace is the third area of the Boost 4.0 framework and is where the Fiat and the Philips trials are focused on. Fiat will connect the AGV with the laser machine in order to further automatize the production process, whereas Philips will improve the control system for the moulding process providing more data to the operator and using data analysis to predict and prevent the defective parts production.

The forth area is Smart Connected Production and is covered by the Gestamp and the Volvo pilots. In the first case, Gestamp will develop a new high-resolution metrology and visualization system to achieve zero-defects manufacturing. For its part, Volvo will work to get a real-time tracking of cabs form Umea to Tuve plants allowing to prevent delays and improve the planning process.

Finally, Whirlpool and Benteler's pilots have as their work area the Smart Maintenance and Service, the first one by updating their forecasting tool for spare parts distribution and manufacturing, including the use of big data produced across all phase of product lifecycle. The second one by developing a machine health monitoring system that will allow to improve the predictive maintenance plan and prevent the defective production.

Keywords: state of the art, present scenario, future scenario, expected results, execution plan

Disclaimer

Acknowledgement

This document is a deliverable of Boost 4.0 project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 780732.

Table of contents

Abbreviations and Acronyms

Acronym	Meaning
A.I.	Artificial Intelligence
ANN	Artificial Neural Network
API	Application Programming Interface
CEP	Complex Event Processing
DSS	Decision Support System
ERP	Enterprise Resource Planning system
FDT	Fault Detection Tool
IoT	Internet of Things
KPIs	Key Performance Indicators
ML	Machine Learning
MES	Manufacturing Execution System
PaaS	Platform as a Service
PMT	Predictive Maintenance Tool
RDBMS	Relational Database Management Systems
SIEM	Security Information and Event Management
SSA	Singular Spectrum Analysis
SMS	Smart Manufacturing Systems

List of Figures

1 Introduction

This is the first deliverable to report the progress of Task 2.3, with the objective of giving an overview of the pilots, describing in a general way each of the pilots and the planification of the experiments, identifying the required technologies adaptation and designing a concrete plan of implementation for each of them.

As the implementation process will be an iterative and incremental process, there will be a second deliverable for this task to report the evolution of the requirements and the pilots.

1.1 Scope and organisation

This document is the starting point for WP4-WP8, as it summarizes the present scenario, the problems that want to be solved and how to address them as well as the expected results.

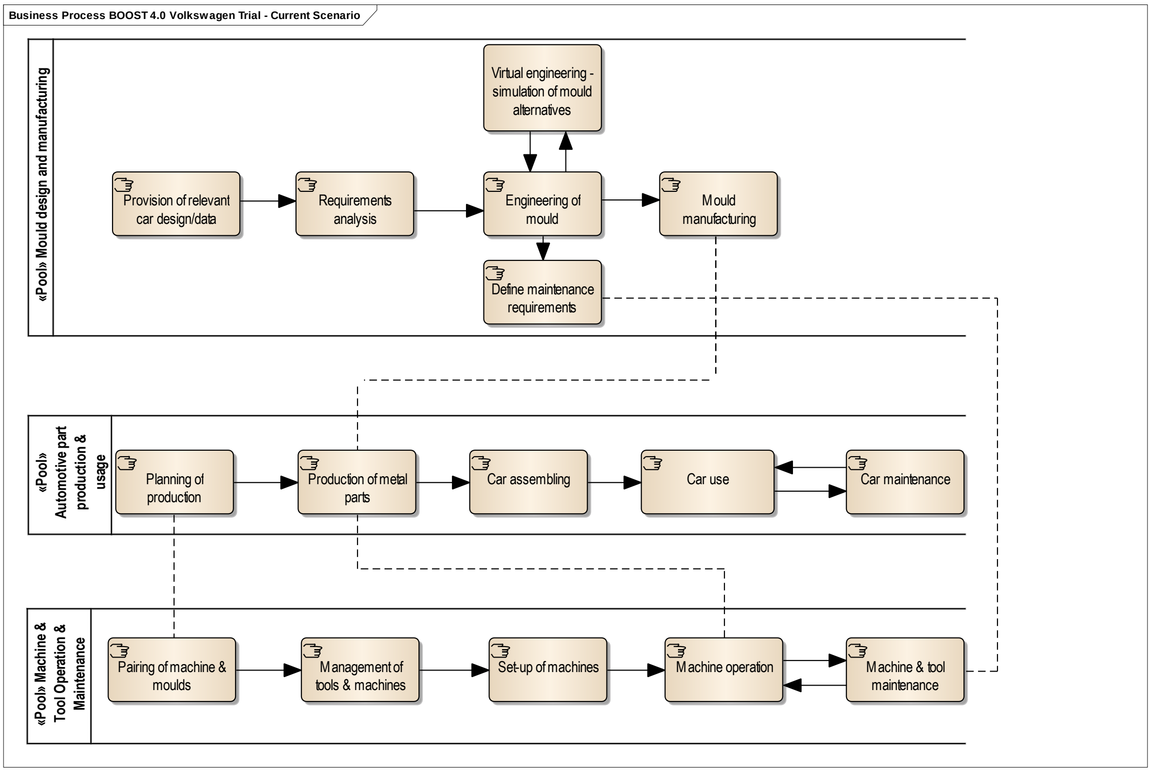
Precisely that's the structure of the document: for each of the 10 pilots there is a section to describe the state of the art of the main technologies and solutions involved, the present scenario, weaknesses and bottlenecks, which is the expected the future scenario after the implementation, the expected results of the experiments and the execution plan for each trial.

2 Trial 1: Volkswagen Injection Moulding Plant

2.1 Trial present scenario

The present scenario of production of light metal castings is a conventional approach. A mould as a tool for the production of specific parts must be designed and manufactured, this tool is paired with a light metal casting machine which is used as a process step in an entire automotive manufacturing process. The VW scenario therefore deals with three main processes in the production of metal components:

- The mould design and the manufacturing of the corresponding tool
- The assembly and use of the component as well as the operation and maintenance of the associated machine in combination with the manufactured tool.
- The production of the light metal components



.

Figure 2-1: BOOST 4.0 Volkswagen Trial - Current Scenario

A conceptual overview of these three main processes, included process steps and their associations are shown in Figure 2-1 and are described in the following.

Mould design and manufacturing – In this process the mould is designed and manufactured. In the first step, the relevant car design data (CAD) is needed to have a 3D model of the part which has to be produced.

The subsequent step covers the requirement analysis. Based on the dimensions of the part, the design, the raw material for the part production, etc., the mould has to fulfil a range of requirements, e.g. cooling & heating times, durability, tolerance, etc. to keep the quality and performance of the production process as well as the metal component. Note, therefore also requirements of the raw material have to be defined. Such requirements are defined in the requirement analysis process step.

After defining the appropriate requirements for the component and the mould, the engineering of the mould can begin. In this step the mould gets engineered in as an CAD model and simulation software supports the engineer in finding a requirement fulfilling mould design. The engineering of the mould and the support through simulation are two alternating process steps.

The finished mould design also defines a range of maintenance requirements to assure a continuous production process which reaches the component quality and production performance.

The final process step is the manufacturing of the mould itself based on the mould design. The manufactured mould is ready for the pairing with the light metal casting machine.

Machine & Tool Operation & Maintenance – This process has the view on the metal casting machine which gets paired with a specific mould and executes the light metal casting process. It contains relevant process steps for the metal casting machine in connection with the manufactured tool. The interface between the tool and the machine must be taken into account and included in the production planning.

The first step is the pairing of machine and moulds. In the machine gets a reconfiguration through the replacement of a mould for a specific component as designed and manufactured in the *mould design and manufacturing* process.

The next step is the management of tools and machines.

The subsequent process step is the set-up of the machines which includes the configuration and ramp-up of the machine so that its integrated in the manufacturing process for automotive part production.

The following machine operation step represents the operation of the machine during the production. It is alternating with the last step “machine and tool maintenance” which represents interruption through a defined maintenance plan for the machine and the mould. The already defined maintenance requirements in the *mould design and manufacturing* process are also considered in this plan.

Automotive part production and usage – The last process represents the production of automotive parts in Volkswagen and the usage of the final product.

The first step represents the planning of production where ERP-based, production, machine and tool requirements are brought together to elaborate the plan for production.

The subsequent step is the production of metal parts as a focused in the Volkswagen pilot. This step represents all necessary actions for the light metal casting process starting from the raw material ingestion an ending with manufacturing light metal component. The component production can start with the aid of the manufactured tool and the relevant machine. The finished metal components are subjected to several quality checks.

The next step is the cars assembling which represents all necessary actions to assemble other components into a functioning automotive.

Finally, the automobile is used and alternates with the car maintenance process step in which temporarily components get maintained.

Relations between the three processes – The described processes steps have a couple of associations. The defined maintenance requirements for the mould directly influence the maintenance of the metal casting machine and tool. The manufactured mould has an effect on the production of metal parts. The production of metal parts is associated with the machine operation. And finally the planning of production decides about the pairing of machine and correct mould.

2.2 Trial future scenario

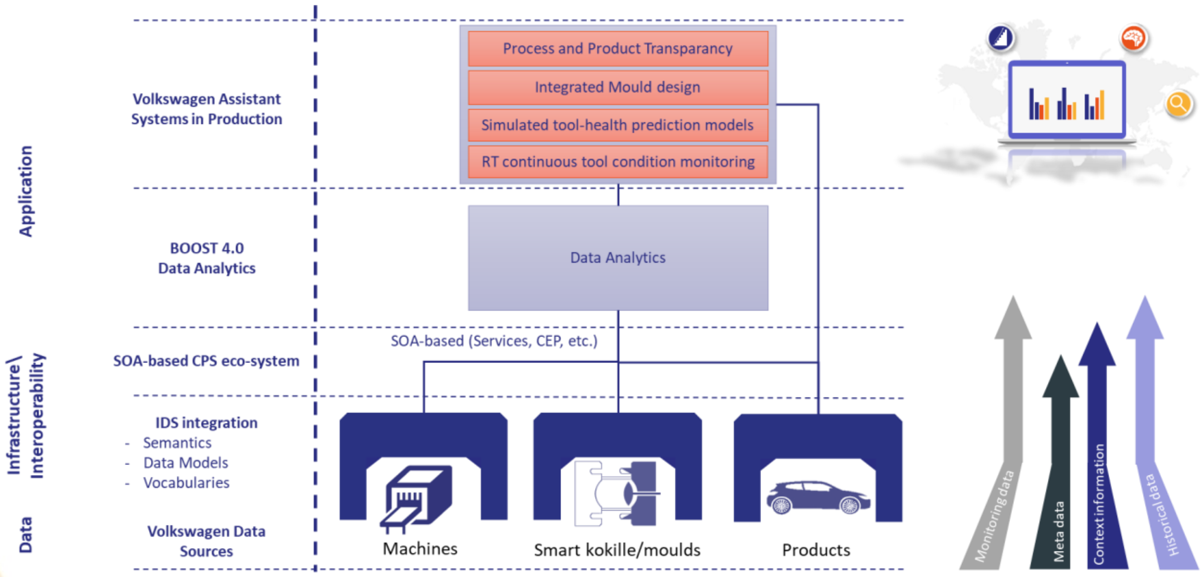


Figure 2-2 BOOST 4.0 Volkswagen trial - conceptual overview

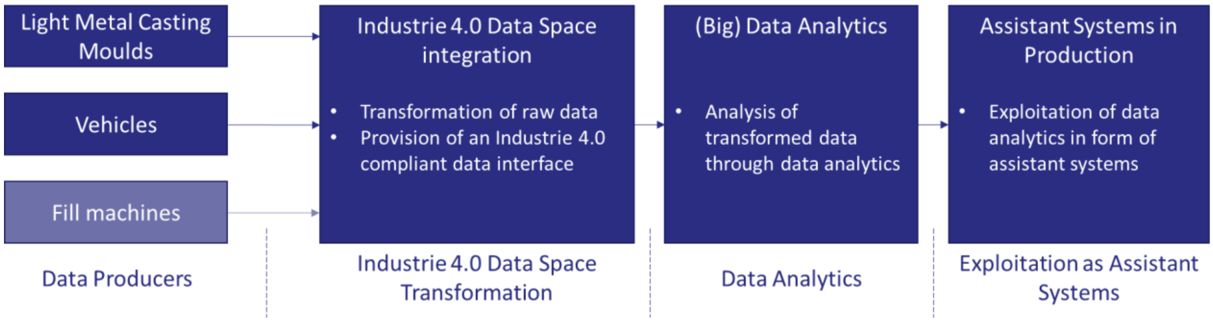


Figure 2-3 Data flow, data processing and data exploitation

As conceptually shown in Figure 2-2 and Figure 2-3, in the future scenario Volkswagen creates new data sources and enables unified access to valuable data produced by products, machines and moulds so that the brownfield light metal casting processes are transformed into an Industrie 4.0 compliant data space. This is achieved through the unification of data protocols and the approach to diverse departments locally and at diverse locations and possibly even transferring related semantical unification to generate common communication protocols/languages (vocabulary, grammar) and allows an integrated approach for modelling information, gathering data, deriving information & knowledge as well as to transfer knowledge to first tier suppliers

This transformation enables the exploitation of available data through direct (raw or pre-processed) data provision or data analytics services which provide their analytics features and results as user friendly assistant systems in the smart manufacturing environment.

The future state in Volkswagen realizes a zero-defects production that allows an evolutionary introduction in heterogeneous manufacturing environments. New assistant systems provide features for real-time continuous tool condition monitoring in the light metal casting process and simulated tool-health prediction models allow predictive maintenance.

Another part of the future state is the feedback to mould design. Analysis of data coming from products, machines and moulds are used to improve the design of moulds and their virtual models used in simulation tools.

The whole future state allows the integration of the approach also in departments that are operating old machines and tools with rather limited abilities for data provision as well as with machines, moulds and tools that are already able to provide rich data sets in real time

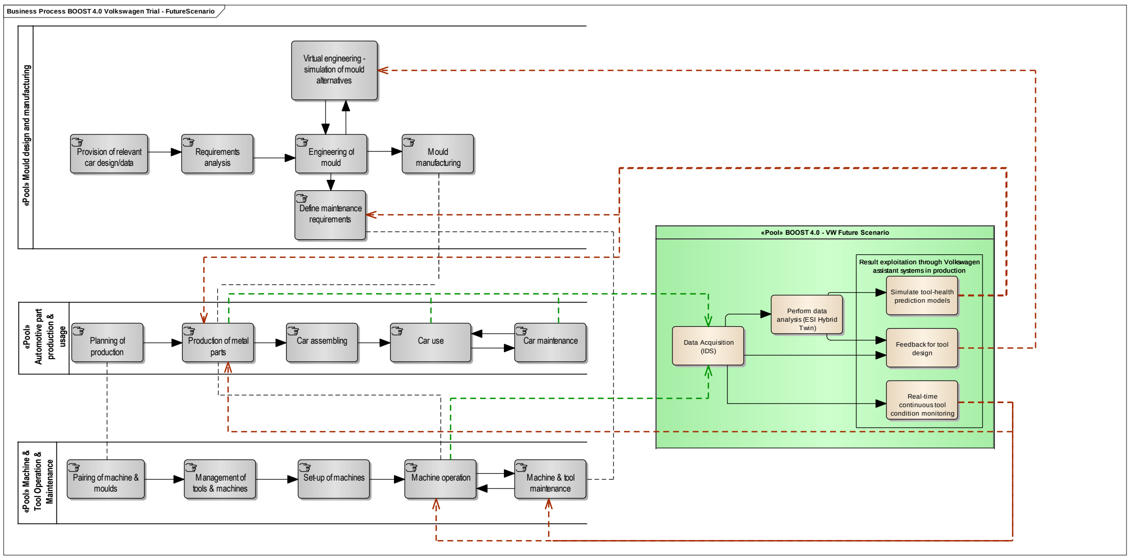


Figure 2-4: BOOST 4.0 Volkswagen Trial - Future Scenario

Figure 2-4 shows the future state in the Volkswagen pilot based on the current state. The figure extends the workflow presented in Figure 2-1 and shows in green boarders the future state as workflow for the Volkswagen pilot. The green lines show the data acquisition of important data in the Volkswagen pilot, machine operation, the light metal casting process, and the vehicle. For these are used the approaches of the industrial data space and their connectors. These data are used directly by Volkswagen production assistant systems before analysed in analysis tools as the ESI hybrid twin which provides analysis results to assistant systems.

Assistant systems in the future scenario are:

- **Simulation of tool health prediction models** – Analysis tools are used to generate tool health prediction models.

- **Feedback for mould design** – Real data coming from the light metal casting process and vehicles in use and enhanced simulation models are providing information to improve the mould design and also the simulation of alternative mould designs.

- **Real-time continuous tool condition monitoring** – Available data are used to for a continuous tool condition monitoring. and tool maintenance.

3 Trial 2: FILL Gurten

3.1 State of the art

3.1.1 State of the art of digitalization

For the digitalization of the value creation network, the trial partner FILL has developed the customized System *Machine Workflow* in the past years, which collects data from the production process of the machine, stores them in database systems and makes them available for further evaluation with regard to the produced product. Communication between companies in the value-added network takes place via web services. The machine workflow controls the material flow in production directly and automatically.

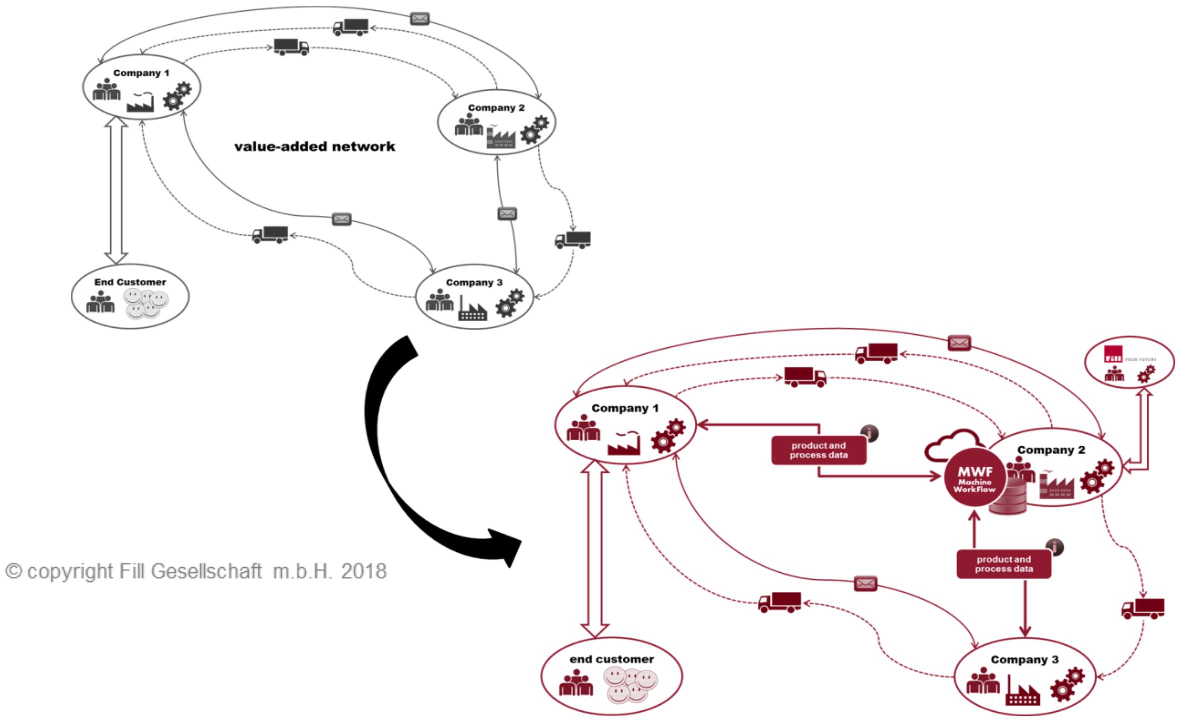


Figure 3-1: Horizontal integration of the machine workflow

3.1.2 State of the art of the engineering process and management

According to the engineering process, the operational and organizational structure at Fill is well designed and accessible to all employees via an online process map. Especially in traditionally conservative mechanical engineering, the engineering management is a heterogeneous grown infrastructure. Best in class tool strategy is the preferred management strategy with all of its advantages like:

1. High efficiency in the individual departments and disciplines
2. Satisfied employees

However, it also has disadvantages like:

1. Creates heterogeneous tool landscape
2. Data exchange requires a great deal of communication and transfer work
3. Inefficient and error-prone processes
4. Automation of internal change management becomes very difficult

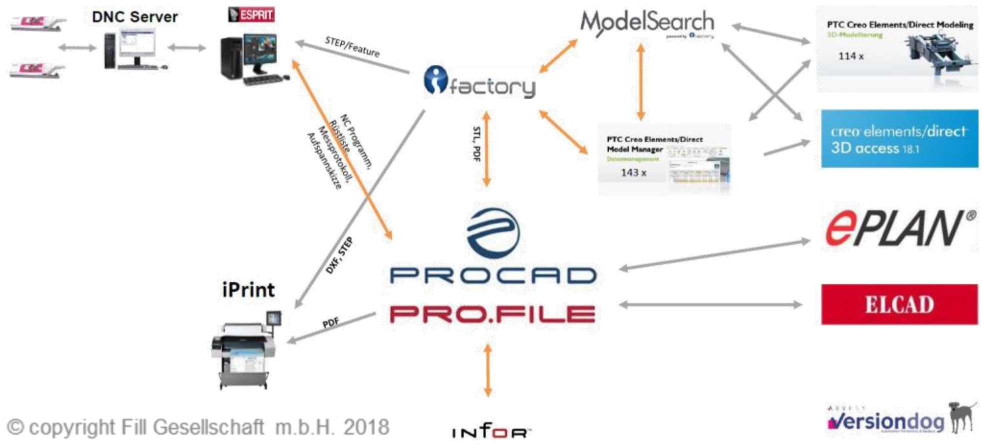


Figure 3-2: present infrastructure

During the engineering process state-of-the-art approaches like modularization, model-based engineering concepts and simulation as well as virtual commissioning are used. This methodology enables simultaneous and agile engineering processes.

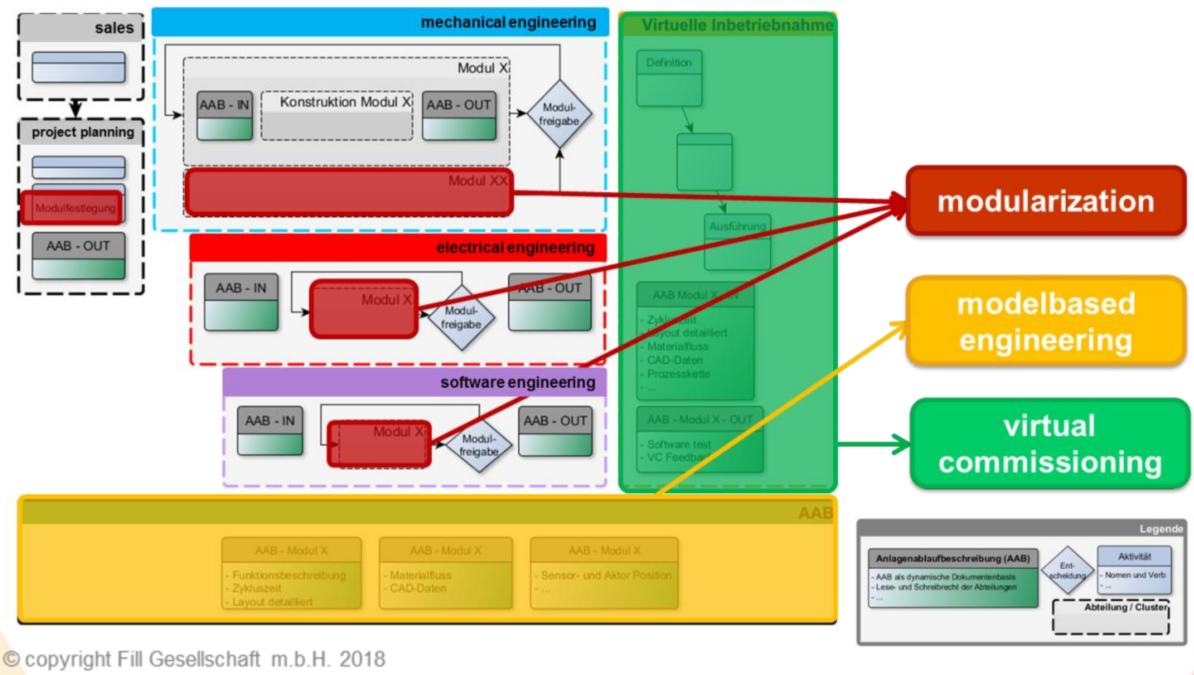


Figure 3-3: state-of-the-art methodology

3.1.3 State of the art and ongoing work of data analytics

Fill and RISC are working on the scientifically close research project VPA 4.0 – Virtual Production Assistant funded by the Government of Upper Austria. The goal of this project is the conceptual design and development of a virtual production assistant (VPA 4.0) based on data and visual analytics of machine and job data, providing a learning-based knowledge base for the machine. To achieve this, all data is imported into a hybrid database system. Big data streams (BigData) are stored in a NoSQL system using standard importers (e.g. OPC UA), while model information of the machines is stored in the form of ontologies in a generic database system. The ontologies are generated by means of a graphical editor as well as further processing and additional data such as, for example, ERP data and machine configurations are stored in the generated data model by means of the ETL process.
The VPA 4.0 is primarily used to assist local experts. It allows a better understanding of the machine by recognizing cause-effect correlations due to anomalies and patterns. In addition, maintenance intervals and cycles can be optimized and as a result, quality improvements of the production and the product can be achieved.

3.1.4 State of the art of 3D simulation and visualization

Visual Components is a powerful, flexible and scalable 3D simulation and visualization platform, which allows to easily creating any factory layout, including automation and robotics, to analyse and evaluate performance. The digital twin created in the simulation mirrors the real systems with allows easily validate the real system even before it has been built. By using virtual commissioning, ramp up times are considerable reduced and downtowns during runtime operation are minimized. [1]

The analysis of the data generated in the simulation during the conceptualization phase allows optimizing the development of the new system design from the beginning of the life. The digital twin will be developed along the entire life of the system allowing to enhance efficiency and productivity.

Visual Components solutions provide several functionalities:

- Easy to use and intuitive UI
- Large amount of pre-existing virtual models, parametric, component modelling.
- CAD import to create any new virtual component
- Layout configuration tools
- Project ready deliverables
- Advance Robotics tools
- Communication interfaces, OPC UA and Web services.
- Open interfaces to interoperate with any platform

3.1.5 Edge Nodes for Industrial Automation

Edge or fog computing [2] can be seen as a complementary concept to cloud computation. Instead of using only powerful (but usually remote) computing clusters, some tasks are performed at the edges of the system. Reasons for this can for example be better scalability, lower bandwidth usage, partial local autonomy or abstraction of local differences. Ideally the use of edge nodes should be largely transparent for end-users. Instead the end-user has a service-level view on the services provided by the system while the system decides which services are provided by which parts of the system. An industrial edge node could for example incorporate one or multiple of the following devices: industrial PCs, industrial network switches, gateways, and analog and digital I/O. The edge node to be used in this project as hardware platform is from Nebbiolo/TTTech and constructed in a modular fashion such that it easily scales in terms of computational resources as well as in networking capacity and I/O capabilities (Figure 3-4). Most importantly, the edge node hardware and software provide industrial-grade non-functional properties in terms of real-time behavior, reliability, availability, and security. In R2C2 these nodes will fulfill multiple roles: as controllers for the underlying software-defined real-time network, as transitions points between automation and IT network and as distributed (real-time) computation and control platforms (e.g. for AI applications) in the overall system architecture.

Figure 3-4. A prototype of an edge node by Nebbiolo Systems/TTTech.

Other industrial edge nodes are for example "TwinCAT IoT" [3] and "netIOT Edge-Gateway" [4] developed by automation companies Beckhoff and Hilscher, respectively. The nodes allow the exchange of data between devices in automation and IT/cloud networks. Historic and near real-time data is routed into dedicated cloud data pools for later retrieval or historic analyses. However, aspects such as hard real-time properties, prioritization of data traffic or networking reliability remain largely neglected.

3.1.6 State of the art of big data and data analytics in manufacturing companies

In the course of Industry 4.0, the manufacturing industry began to equip machines with sensors. For this purpose, protocols were developed and standards extended or improved in order to transmit the information from the resulting sensor networks in a standardized manner. OPC Unified Architecture (OPC UA) as a machine communication protocol has been defined as an industry standard by the OPC Foundation as a successor to existing COM-based standards. This made the data connection to a machine network possible, in which way large amounts of data are available. These efforts gradually allow the use of OPC UA in the manufacturing industry [5]. The data sets can now be analysed using methods from data mining, such as machine learning, which can be used to derive information about correlations existing patterns and anomalies in the data pool.

From an ICT point of view, the basis for predictive maintenance and optimized set-up times is the use of statistical methods to analyse data. As a result, patterns can be identified and subsequently derived mathematical models. The combination of sufficiently meaningful information and correct procedures forms the solution for a suitable model.

Today, deep learning / machine learning methods have been successfully used in image processing, speech and language recognition (e.g., computer linguistics) (e.g., Cortana, Siri), as well as text analysis and automatic translation (e.g., Skype). Deep learning is a branch of machine learning based on artificial neural networks to represent and (parallel) process data at different processing levels with complex structures. In this case, one speaks of "deep neural networks" [6]. which represent the input level on an output level through several "hidden" levels. The following prerequisites recommend the use of deep learning: large data volumes, many connected input properties or multitask settings. Deep learning requires a lot of computing power, so there is a big challenge in the efficient implementation.

3.1.7 State of the art in the use of ontologies

In computer science, the use of ontologies is a common methodology for modelling knowledge digitally and formally. The term ontology originally comes from the field of theoretical philosophy. In computer science, an ontology is a formal description of complex facts and expertise (to conserve and share it, but also to machine it (further) to process and expand [5] [7] [8], such as the genes ontology [9] or Disease Ontolgoy [10]. While the use of ontologies has long been limited to the scientific-academic field (especially computer science, biomedicine or biotechnology and medical informatics), in recent years more and more applications in the area of big data, data analysis, data integration and industry 4.0 have been found [11]. Even Industry 4.0's Industry Reference Architecture Model [12] notes that semantic technologies (aka ontologies) will make a significant contribution to the successful implementation of Industry 4.0.

An ontology [13] uses semantic search technology to discover meaningful information from structured and unstructured data. With their products (e.g., Ontology Intelligent 360 for network operators) and platforms (e.g., Ontology 5), they address the challenges that data integration presents (data sourcing, data connectivity, data migration). They want to forego in-depth MDM, ETL, BI, or data warehousing integration projects, develop their products based on a semantic data model that requires no schema, and built using the Graphene-based technology. Ontologies overcome the lack of flexibility and the significant delay time that data integration cannot cope with during changes in data types, data sources, or datasets. A big topic for ontologies is the connection of data from different data sources and their interpretations [14]. They argue that the data agility, data-first approach - data-building (data first, schema / structure later) - and the step-by-step data integration process provide a foundation for a new way to manage enterprise data collections and BigData [15]. They develop products that can model the complex networks of things. One example is the Ontology's Common Cause Analysis (CCA) module, which was developed for Telenor Denmark to perform a rapid analysis of network failures and provide a solution to the root causes [16]. In addition, ontologies are used to structure BigData adequately. To read is a process under Gavrilova et al. [17].

3.1.8 State of the art in structuring BigData using ontologies

A notable work in the context of this project is the article Big Data Semantics in Industry 4.0 [18] from the year 2015. In the authors describe the need to enrich the numerous and extensive sensor data with semantic information in order to draw meaningful conclusions. To describe the sensors, they resort to a ready-made ontology [19] describing sensors and their data. The authors limit their work to sensor data and (unlike the research project here) do not include any other data (ERP data, parts lists, CRM data, etc.).

The work of Grangel-González et al. Is most relevant for the current project. titled Towards a Semantic Administrative Shell for Industry 4.0 Components [20]. The authors describe how they want to implement a so-called administrative shell with the help of semantic technologies (in this case with RDF - Resource Description Framework). A concept that comes very close to the virtual production assistant aimed for here. An administration shell holds data from various producers (machines, hardware, software, etc.) as well as meta-data and semantics and makes them available. The concept of the administration shell is described in more detail under [21].

A central aspect of this project is the involvement of the experts in data analysis and interpretation. Ontology-based solutions that achieve this in the complex field of medical research have already been successfully published **by the applicant** ***RISC*** [22] [23] [24]. Gorecky et al. In their work [25], they deal extensively with the changed role model of the expert in industry and identify ontologies as a means of semantic support in data interpretation.

Partner	Technology	Description
FILL	Engineering processes & engineering mgt.; Logging & monitoring	• Engineering of highly automated production systems • Development of engineering methodologies • Standard fieldbus connectivity between fieldbus participants like PLC • Machine data logging with OPC-Technology • Implementing performance indicator systems like OEE metrics
RISC	Big data and data analytics	• Statistical methods as well as methods from machine / deep learning • Data and visual analytics of machine and job data • Development of a learning-based knowledge base for the machine • Combination of data- and model-driven approaches in order to create more promisingly results
TTTech	Edge/Fog Computing Platform	• Edge device platform with supporting fieldbus connectivity and switch functionality for standard deterministic ethernet connectivity.

		• Software stack that provides a secure virtualized platform for virtual machines, containers and application hosting • Centralized management system for updating device software and deploying applications
VIS	3D Simulation and Visualization. Digital Twin and Virtual Commissioning	• 3D simulation and visualization platform, which allows to create any factory layout combining logistics flows and automation. • Reusable along the entire factory life cycle from conceptualization until decommission. • Scalable from a simple machine to the entire factory. • Open interfaces to interoperate with any platform • Communication interfaces, OPC UA and Web Services

3.2 Trial present scenario

With the FILL trial it is possible to record the data of their machines standardized by OPC UA and subsequently to use them for analyses and optimizations. With a standardized communication technology, the existing specific solution Machine-Work-Flow-Framework can be generalized and used for further customer requests. In doing so, FILL takes a big step forward in the digitization of its machines with the expanded machine state model. FILL pursues the following goals:

- Cost reduction expected by reducing the time spent on future development and customer projects.
- Development of data-driven business models in service and support
- Identification of optimization potentials in the engineering process for a long-term reduction of the development times of machines.

The FILL trial primarily serves the engineering process of the machine builder. It allows for a better understanding of machinery by detecting cause-and-effect relationships due to anomalies and patterns. In addition, maintenance intervals and cycles can be optimized and, as a result, quality improvements of the production and the product can be achieved. Currently there is a lack of digitalisation mostly through manual information flow (see Figure 3-5). It is planned to resolve this lack within the present pilot.

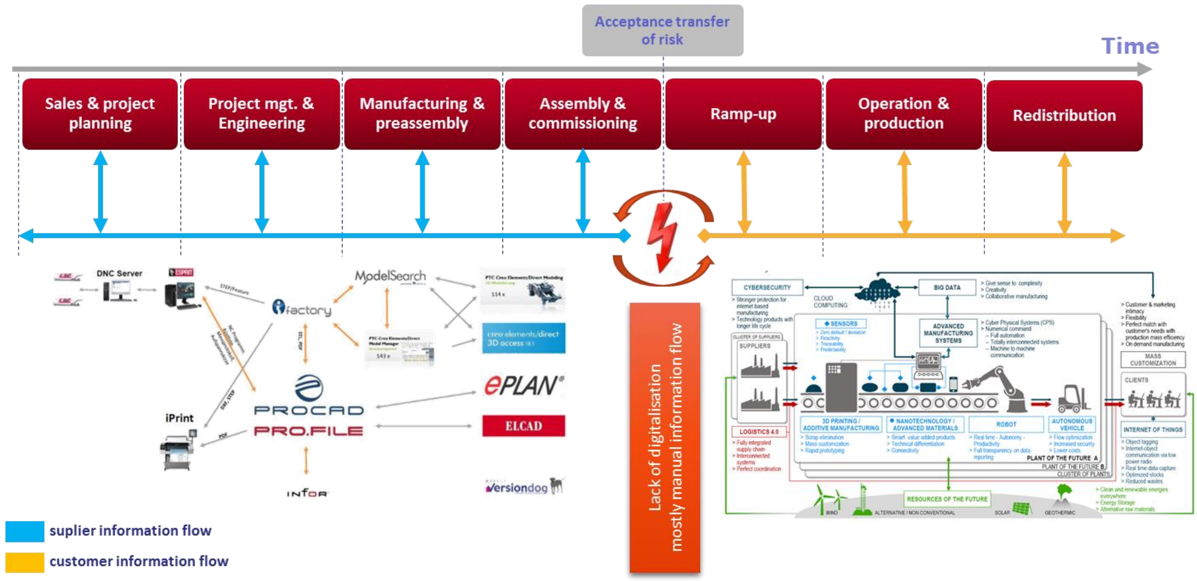


Figure 3-5: production system life cycle and information flow

The technological possibilities for measuring, recording, storing and processing machine and production data are constantly improving. In individual areas, e.g. the CAD/CAM process between engineering and production, systems with feedback loops (closed-loop system) are created. Due to media discontinuities, manual evaluations and activities, however, the existing systems are mostly open-loop systems. This means, especially for globally (distributed) systems, that the response to faults or changes can only be inadequate, delayed and outdated. In order to achieve the vision of a digital, intelligent, agile and self-controlling value-added network (engineering and production), the information loop must be closed across all systems involved.

Actually, the information flows between customer and supplier are very diverse. Most of the time these are informal information channels, e.g. the operator (customer) calls the engineer (supplier) directly. Getting the customer information into the engineering management system is therefore a person-dependent process so that the requirements management as well as the tech. change management are insufficiently supported. There are also support and service information used during the engineering. Data analyses to gain knowledge for new developments are currently not carried out. Therefore, the lead-time and quality of new developments depends primarily on the experience and communication skills of the employees in the project team.

3.3 Weaknesses and bottlenecks

The FILL trial has its origin for the users in the combination of the following four initiatives and therefore has a high risk for the implementation:

- Maximize energy efficiency (operational level)
- Flexibility and cost efficiency (tactical level)

- System solutions (strategic level)
- Industry 4.0

Technical feasibility is at high risk in terms of data quality and performance, as large amounts of continuous data from different data sources and systems have to be processed in real time.

One of the biggest obstacles to the use of machine learning and deep learning approaches is the lack of adequate training data. Furthermore, the chosen methodological approaches to quality monitoring can lead to results that are theoretically possible, but the practical suitability cannot be proven and therefore does not provide any benefit.

The following problems can currently be identified:

Partner	Weakness and bottlenecks	Impact to FILL pilot
FILL	Historical grown IT architecture makes it difficult to implement new (standardized) interfaces. Therefore, knowledge of special experts is necessary.	The possibility that there is only one expert for different topics and the dependences on software and IT providers can lead to delays
FILL, RISC	Manual information transfer causes loss of information and makes standardization more difficult.	All information is needed for creating additional knowledge through data analysis. Therefore, high effort is put into digitizing all information.
RISC	Data quality is insufficient or data streams are incomplete.	Specific data engineering can be necessary before starting with data analytics.
FILL, TTTech	A high diversity of interfaces, data structures and protocols is given.	Simplified and vendor independent data integration platform for data aggregation will break the brand silo that exists today.
FILL, TTTech	Suitability of existing data management systems is used as an integration platform for distributed data sources, e.g. IDS.	A unified interface and a common language for M2M (e.g. OPC-UA over TSN) is the key enabler for Big Data analytics in manufacturing industry.
FILL	Referencing and linking of engineering, machine and production data is done on machine builder side (supplier side). A data-feedback-loop from the customers is not taken into account.	This feedback-loop is necessary to create a better understanding of the machine; therefore, close cooperation with the pilot partners Benteler and VW is in progress.
FILL, RISC	Engineering processes are not designed for data driven approaches.	Using the paradigm of the engineer-in-the-loop, **knowledge-creation for the engineering process is the goal of the FILL pilot!**
FILL, RISC	Existing engineering data management system is not designed for globally	A strategy is developed to make optimal use of the industrial data space.

	distributed data (e.g. no cloud technology involved).	
FILL	Customer expertise is missing – lack of acceptance.	close cooperation with the pilot partners Benteler and VW is in progress
FILL	Data governance between machine builder and customer is not clear.	A strategy is developed how to clarify the data governance issue (based on results of WP2-T2.5)
VIS, FILL	CAD data sets in the FILL workflow cannot be directly used in VIS to create the digital components	A strategy is developed to use in the workflow a middleware format based on ISO 10303-21
VIS	Limitation in the amount of data generated in the simulation models to be used by the other partners	It would be necessary to re-engineer the virtual models used in the simulation and create new architecture to the requirements of the other partners
VIS	Communication overheads due to the amount of data generated in the simulation during runtime	It would require to redesign the communication architecture and introduce new interfaces

3.4 Trial future scenario

<u>Advantages of the future scenario:</u>

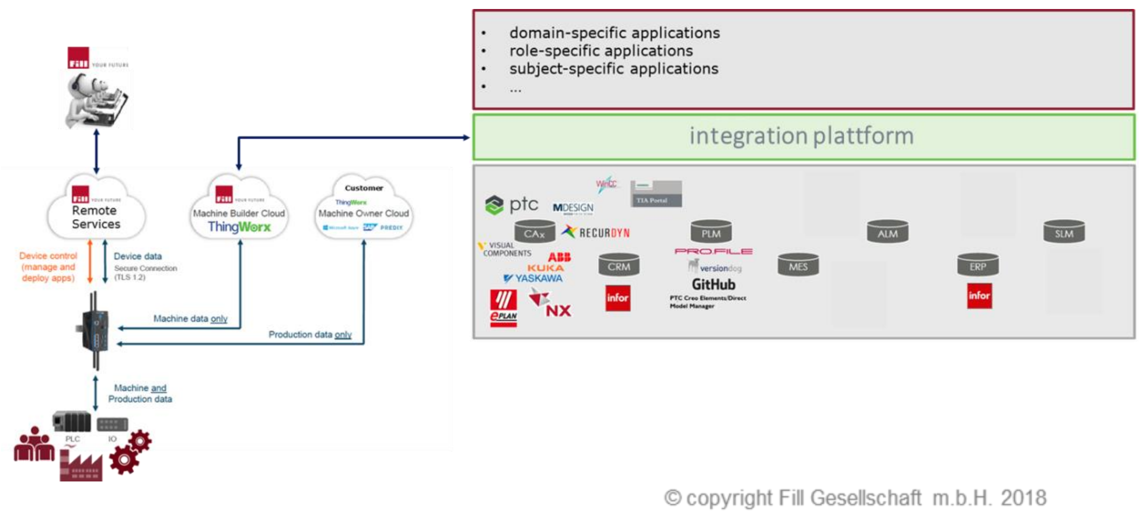


Figure 3-6: data management and communication concept (overview)

Reduce hardware spending by 50% or more by utilizing virtualization technology for the industrial world. Sharing computation resources between different operating systems and applications enables to the convergence of various functionality (e.g. industrial PC, gateway, PLC and firewall) in one device. By enabling resources to be virtualized, the approach manages the hardware work for the user and/or operator of the plant.

Seaming less **data connectivity** enables data to be used in three ways: at the edge, on the local server or remotely in a cloud. This gives machine builders and plant operators the flexibility to choose where and when data is stored, visualized and processed. Data storage is handled by a time-series database hosted on fogOS at the device level, or on fogSM hosted locally on a server or remotely in a cloud. Data transport between device and system manager is configurable, providing an easy to use, consistent view of data locally and centrally. fogOS supports multiple standard interfaces (AMQP, MQTT, REST/JSON, OPC UA), meaning data can be integrated quickly and easily into existing IoT cloud solutions such as Microsoft Azure, SAP, IBM Cloud, AWS and more.

Open interoperability as the platform is designed with openness, interoperability and flexibility in mind. The platform helps to become vendor independent, giving the user the freedom to choose the solutions, from any provider, which best suit the specific business needs. The Linux-based fogOS provides a standard application environment, meaning that there is no need for developers to work on the software infrastructure around apps. Python SDKs with REST APIs are available to aid integration.

Provide flexible support and cut maintenance costs since the platform is envisaged to provide a secure deployment mechanism with signed software verification at the device and transfer over encrypted TLS (SSL) connections. Updating software from fogSM significantly reduces maintenance time and effort and helps to ensure consistency of software in devices. This can be used to respond to security threats or critical bugs, where patches can be applied quickly and accurately. The fogSM application store can also be used for deploying new applications to add functionality to multiple devices simultaneously or to selected edge/fog nodes only.

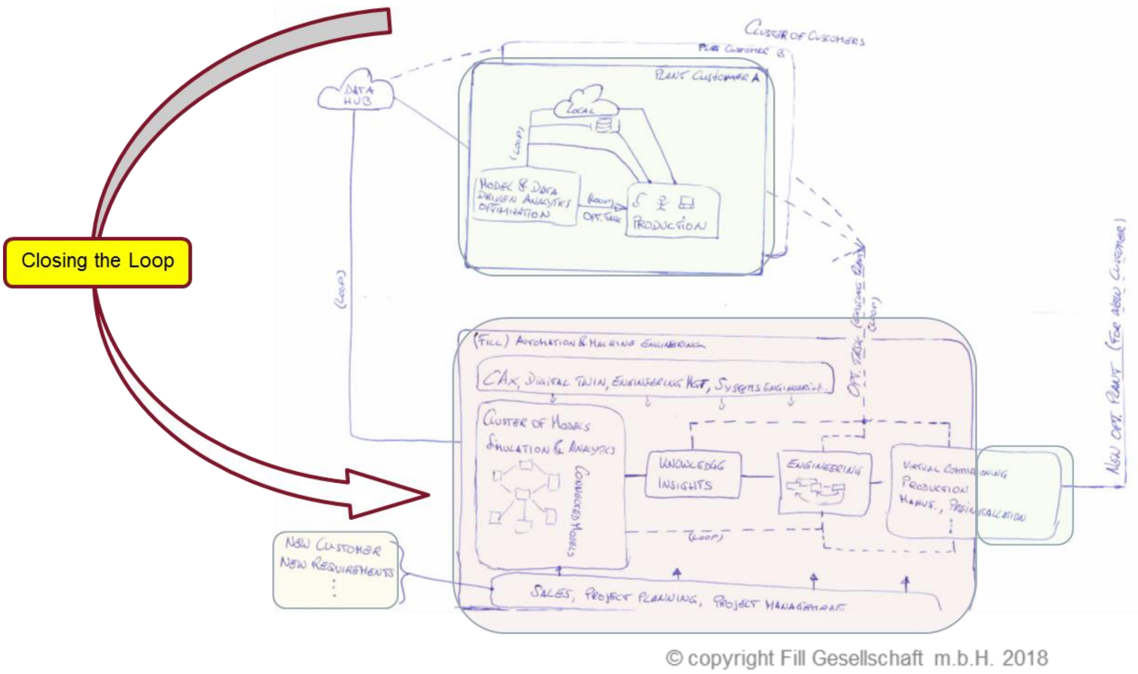


Figure 3-7: Big Picture

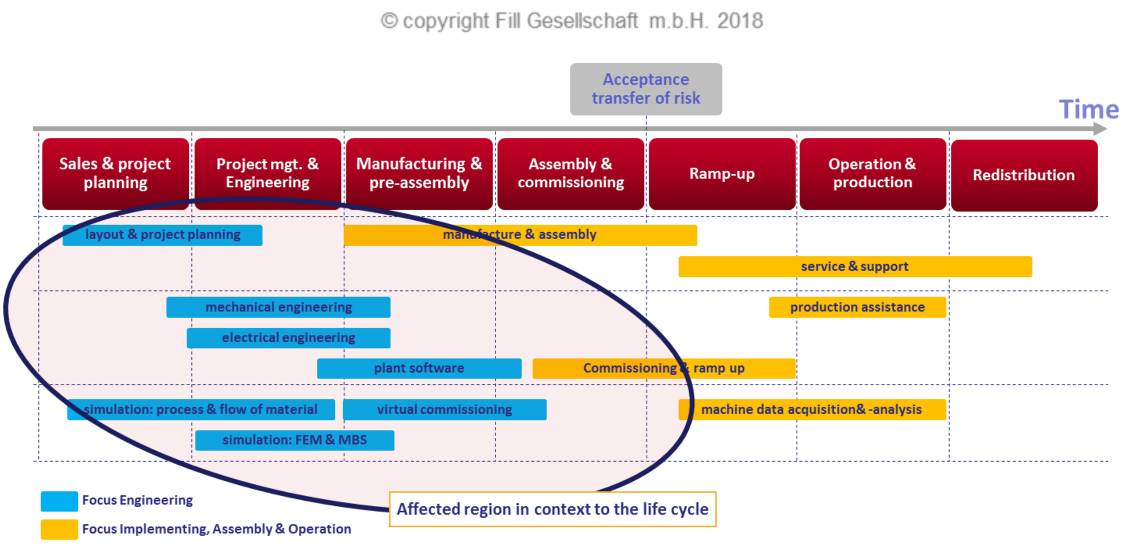


Figure 3-8: affected regions in context to the life cycle

3.5 Expected results

3.5.1 Expected results for FILL

Maintenance technicians of production machine must guarantee high machine availability on a daily basis while reducing the cost of maintenance and spare parts storage. With the increasing networking of production systems and the systematic use of machine data, the possibilities of digitization in the context of Industry 4.0 open up new opportunities. Not only for the trial leader FILL, but for all machine manufacturers and users, the production machines are currently the same as a black box during operation. The state of the process and quality-critical machine components are hardly or not at all known to the machine operators in terms of their granularity. It is worse with regard to the machine builders (like FILL). Using knowledge from the operation phase for new machines is only possible through a high communication effort.

3.5.2 Expected results for RISC

In the field of data analysis and knowledge generation, the use of ontologies has so far been mainly used in medical research, and in recent years has increasingly been used in production and logistics. There, however, the volume of data is much higher and cannot be processed. The combination of ontologies on the one hand and Big Data on the other hand is an enormously important field in research as well as practice. By extending the sphere of influence to the so-called Big Data range, this system idea can be used successfully in the long term also in the industry. Conversely, existing ontology-based analytics (visual analytics, visual clustering, principal component analysis, etc.) can be applied to Big Data, which in turn brings great value to existing and new applications.

Most of the IoT and Big Data platform approaches focus on operations and not on the engineering process of such production systems. Closing the loop leads to the next evolutionary step in the engineering methodologies like model-based engineering and model-based systems engineering. Lossless and highest standardized interoperability between engineering tools, processes and domains enable the automation of communication work and provide the right data to the right destination at the right time.

3.5.3 Expected results for VIS

Participating in BOOST 4.0 and particularly in the FILL pilot will boost the development of Visual Components' simulation family extending its functionality for handling big data. In addition, the communication interfaces will be extended, particularly OPC UA for handling big data. This will require the development of a new architecture and the extension of the interfaces for managing (next to) real time communication.

The data serialization and interoperability with other systems will be developed within the pilot, providing support for AutomationML, developing and implementing the storage and exchange of plant data engineering information. In addition, it is expected to develop PLM support through a designed PLM interface with add-ons for PTC Windchild and Siemens Teamcenter.

From today's perspective are the measurable KPIs the following:

Partner	KPI	Measurement
FILL	lot-size-1 engineering lead time	The new engineering process leads to lead-time improvement. After the development of a standard measurement method for measuring and comparing engineering lead times for height customized productions systems. One comparable project should be measured until the end of the BOOST 4.0 project. As result, **lead-time reduction of min. 10% is expected**.
FILL	Human resource development	Developing a recruiting strategy to reach and attracting the best talents from the ICT Community.
FILL	Offering master and bachelor theses	Two Theses per year in the field of IoT-Architect and Data Scientist
FILL	PhD thesis in the field of behavior modeling of machine components	A PhD student with background of physics is starting in autumn 2018.
RISC	Staff development: data scientists	For the research focus Data Analysis (as in the project Boost 4.0), it will be necessary to recruit and build up further Data Scientists in the next few years. One Data Scientist per year is planned.
RISC	Staff development: Big Data Architecture	Due to the big data problems, experts are indispensable for data architectures of large data volumes. RISC Software GmbH plans to expand its existing team in the next few years with another colleague per year.

RISC	Number of international partners	As an Austrian research company the international view is limited. Through the project Boost 4.0, RISC Software has the possibility extend the network of international partners and gain further international (research) projects.
RISC	PhD in Artificial Intelligence	Torsten Welsch is starting his PhD in AI in autumn 2018 with the topics concerning the FILL pilot.
VIS	Increase of sales	VIS expects further customer projects due to the new market opportunities that will be opened up by Boost 4.0.
VIS	Increase of resources of R&D	Further research in the field of unified infrastructure for simulation, communication and execution of robotic systems is planned within Boost 4.0.
VIS	Deployment of open standards	Further open standards defined in Boost 4.0 are being considered to deploy as extension of existing ones such as AML
VIS	Faster product development	The functionalities introduced during Boost 4.0 will reduce the development and commissioning of new machines and workstations.

3.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 1: Smart digital engineering process using industrial big data – closing the loop validated knowledge engineering					
BUSINESS PROCESS 1 - Agile model management & development process					
1	Digital asset repository				
2	Big Data (IoT)				
BUSINESS PROCESS 2 - Data analytics process					
1	Data processing pipelines				
2	Data exploration, model integration and deployment				

BUSINESS PROCESS 3 - Service development process					
1	Extending Fill business model by Digital services (Model as a Service)				
2	Smart Maintenance				
BUSINESS PROCESS 4 - Simulation based release process					
1	IoT Data and simulation driven engineering (using real historic data)				
2	Virtual Commissioning				

4 Trial 3: VWAE real-time self-learning virtual factory 4.0

4.1 State of the art

4.1.1 Big data Interoperability

Enterprise entities are known to own valuable datasets that are used in their daily activities. This data is typically organized, stored and managed in ways that are optimal for their respective internal processes. When trying to leverage the richness of these heterogeneous data sources, by attempting to extract more complex information from the combination of one or more datasets - whether from a partnership between separate enterprises or different internal processes -, an understandable common ground must be reached.

Traditional approaches used to rely on "hand-crafted" data conversion solutions that would attempt to solve each individual problem separately. This would result in ineffective, cumbersome and often redundant software stacks that could not be generally repurposed [26]. In recent years steps have been taken towards a more robust and general interpretation through the application of model-driven approaches to interoperability [27] [28]. The use of model-driven approaches has given enterprises the ability to produce generalized models of their data sources that can be reutilized in multiple solutions.

Having a defined general model for data sources has enabled the use of specialized vocabularies. In particular, ontologies have gained traction as a means to describe particular domains [29] [30] [31] [32]. The Web Ontology Language (OWL) appears as the de facto standard for ontologies and its direct semantics [33], a subset of first-order logic, enrich vocabularies with concrete meaning, allowing not only semantic control over data, but also the possibility of reasoning with it. As a method to describe linked data and as part of the building blocks of the Semantic Web [34], it furthermore facilitates the integration with other linked data sources. Research into semantic interoperability has shown the viability of such approaches within the enterprise context [35] [36].

A recent trend in data analysis shows the rise in the use of high-throughput processing techniques (collectively referred to as Big Data) that leverage the massive amount of raw data typically generated by the enterprise in order to aid in business decisions. Gathering data from heterogenous sources and manipulating them to prepare for big data analysis processes is still a big problem [37]. The BigDataEurope (BDE) [38] framework provides an effort into democratizing the access to big data analysis by reducing the complexity of development and deployment of a big data stack. BDE compares favorably with existing solutions by providing easy extensibility - BigTop, Cloudera, MapR offer no extensibility and Hortonworks has a difficult implementation - while proving an open and limitation free platform. Finally, addressing the need for common information sharing infrastructure, BDE has initiatives concerned with the standardization of data brokering, such as the International Data Space (IDS) [39]. IDS provides an architectural solution through the definition of a concise broker specification that facilitates interoperability among IDS adhering partners.

4.1.2 Big Data Analytics

Big Data Analytics refers to the implementation of analytic tools and technologies within the scope of Big Data. Hence, Big Data Analytics may be described by two specific concepts, Big Data + Analytics, and the interactions between technologies supporting both concepts. On one hand, the concept of Big Data was thoroughly described, discussed and formalized through the three (or more) V's: Velocity (speed), Volume (size) and Variety (heterogeneity). On the other hand, data analytics concept is broadly known, both in academic and business sectors, and comprise Data Mining, Machine Learning, Statistics and other data analysis tools and techniques.

So, why merge these concepts [40]? First, Big Data provides gigantic statistical samples, which enhance analytic tool results. In fact, the general rule is that the larger the data sample, the more accurate are the statistics and other products of the analysis. Second, analytic tools and databases can now handle big data, ad can also execute big queries and parse tables in record time. Moreover, due to a precipitous drop in the cost of data storage and processing bandwidth, the economics of analytics is now more embraceable than ever. Most modern tools and techniques for advanced analytics and big data are very tolerant of raw source data, with its transactional schema, non-standard data, and poor-quality data. That is beneficial because discovery and predictive analytics depend on lots of details—even questionable data. Finally, analytics based on large data samples reveals and leverages business change. The recession has accelerated the already quickening pace of business. The recovery, though welcome, brings even more change. In fact, the average business has changed beyond all recognition because of the recent economic recession and recovery. The change has not gone unnoticed. Businesspeople now share a wholesale recognition that they must explore change just to understand the new state of the business.

4.1.2.1 Analytics applied to Big Data

Data Analytics may correspond to the application of tools and techniques to extract insights and knowledge from data, by analysing it through any of Statistics, Data Mining and Machine Learning techniques. Although statistical analytics is supported by well-known statistical techniques, which are more easily deployed on a Big Data context, in the case of Data Mining and Machine Learning, the passage to a Big Data environment is not a trivial task, since it comprises the reconfiguration of algorithms to be deployed in Big Data execution engines.

In typical data mining systems, the mining procedures require computational intensive computing units for data analysis and comparisons. A computing platform is, therefore, needed to have efficient access to, at least, two types of resources: data and computing processors. For small scale data mining tasks, a single desktop computer, which contains hard disk and CPU processors, is sufficient to fulfil the data mining goals. Indeed, many data mining algorithm are designed for this type of problem settings. For medium scale data mining tasks, data are typically large (and possibly distributed) and cannot be fit into the main memory. Common solutions are to rely on parallel computing [41], [42] or collective mining [43] to sample and aggregate data from different sources and then use parallel computing programming (such as the Message Passing Interface) to carry out the mining process.

For Big Data mining, because data scale is far beyond the capacity that a single personal computer (PC) can handle, a typical Big Data processing framework will rely on cluster computers with a high-performance computing platform, with a data mining task being deployed by running some parallel programming tools, such as MapReduce or Enterprise Control Language (ECL), on a large number of computing nodes (i.e., clusters). The role of the software component is to make sure that a single data mining task, such as finding the best match of a query from a database with billions of records, is split into many small tasks each of which is running on one or multiple computing nodes [44].

Currently, Big Data processing mainly depends on parallel programming models like MapReduce, as well as providing a cloud computing platform of Big Data services for the public. Data mining algorithms usually need to scan through the training data for obtaining the statistics to solve or optimize model parameters. It calls for intensive computing to access the large-scale data frequently. To improve the efficiency of algorithms, Chu et al. [45] proposed a general-purpose parallel method, which is applicable to a large number of machine learning algorithms based on the simple MapReduce programming model on multicore processors. Ten classical data mining algorithms are realized in the framework, including locally weighted linear regression, k-Means, logistic regression, naive Bayes, linear support vector machines, the independent variable analysis, Gaussian discriminant analysis, expectation maximization, and back-propagation neural networks.

Ranger et al. [46] proposed a MapReduce-based application programming interface Phoenix, which supports parallel programming in the environment of multicore and multi-processor systems, and realized three data mining algorithms including k-Means, principal component analysis, and linear regression. Gillick et al. [47] improved the MapReduce's implementation mechanism in Hadoop, evaluated the algorithms' performance of single-pass learning, iterative learning, and query-based learning in the MapReduce framework, studied data sharing between computing nodes involved in parallel learning algorithms, distributed data storage, and then showed that the MapReduce mechanisms suitable for large-scale data mining by testing series of standard data mining tasks on medium-size clusters. Ghoting et al. [48] proposed Hadoop-ML, on which developers can easily build task-parallel or data-parallel machine learning and data mining algorithms on program blocks under the language runtime environment. On the other hand, some authors argued that Big Data Analytics could not rely on existing traditional algorithms, and needed a new set of algorithms which were developed on top of Big Data engines [49].

Other tools are already available for Bid Data Analytics [50]. One of them is Apache Mahout [51], which is a machine learning library built on top of Hadoop. It has MapReduce implementations of major machine learning models, which can help to mine large volumes of data through distributed processing on a Hadoop cluster. Also, Apache Spark has a Machine Learning library called MLlib [52], which includes:

- *Classification: logistic regression, naive Bayes;*
- *Regression: generalized linear regression, survival regression;*
- *Decision trees, random forests, and gradient-boosted trees;*
- *Recommendation: alternating least squares (ALS);*
- *Clustering: K-means, Gaussian mixtures (GMMs);*
- *Topic modelling: latent Dirichlet allocation (LDA);*
- *Frequent item sets, association rules, and sequential pattern mining;*

UNINOVA has a strong background in Big Data Analytics, spanning all its phases and areas. Several works were produced within EC research projects, namely H2020 projects, such as OPTIMUM[1], AQUASMART[2], C2NET[3] or BIG DATA OCEAN[4]. These works were then presented as published papers in relevant conferences, journals and books. Specifically, the Big Data Analytics-related areas in which UNINOVA worked are:

- *Big Data collection and storage for analytics* [53], [54]*;*
- *Big Data harmonization* [54], [55]*;*
- *Big Data Analytics* [56], [57], [58], [59]*;*
- *Big Data visualization* [60]*;*

Furthermore, UNINOVA is working on a novel Big Data software infrastructure, based on Docker containers, and supporting several Big Data technologies, such as the ones presented in this section, in order to enable swift installation and configuration of Big Data clusters, independently of the clusters' specifications.

[1] https://cordis.europa.eu/project/rcn/193380_en.html
[2] https://cordis.europa.eu/project/rcn/194237_en.html
[3] https://cordis.europa.eu/project/rcn/193440_en.html
[4] https://cordis.europa.eu/project/rcn/205983_en.html

4.1.3 AGVs technologies

4.1.3.1 Automatic guided vehicles

An automated guided vehicle (AGV) is a driverless transport system used for horizontal movement of materials. [61]. AGVs were introduced in 1955 [62]. The use of AGVs has grown enormously since their introduction. The number of areas of application and variation in types has increased significantly. AGVs can be used in inside and outside environments, such as manufacturing, distribution, trans-shipment and (external) transportation areas. At manufacturing areas, AGVs are used to transport all types of materials related to the manufacturing process. Only in Europe the last year more than 3000 new AGVs were manufactured [63].

The transportation task of AGVS requires efficient and intelligent routing. Most of navigation systems sold currently can compute these aspects in very short time. Usually these systems are also capable of handling priorities and time schedules. Static routing is a well-established standard in navigation. This routing technology is based on fixed course sections. AGV-Systems with static routing are like the railway system with course sections corresponding to tracks and the central navigation system relating to the railway control centre. Sections are marked as occupied whenever a vehicle is entering and remain blocked until the vehicle has left again [64]. These railways may be really represented on the floor by magnetic tapes, electromagnetic wires, and painted lines, or they may be virtual like in the case of Lidar-based navigation technologies.

4.1.3.2 Communication protocols

The recent convergence between industry and the advanced computing, analytics, low-cost sensing and the practically unlimited connectivity of Internet, is very promising [65]. The innovation level is so high to deserve the name of “revolution” in terms of production speed and efficiency. The new concepts are all around the “data” and the amount of data that can now be collected from industrial plants and machines thanks to embedded sensors and instruments. Plant, machine, and overall system performance can be improved dealing and analysing these data by means of new services [66]. As part of this new trend, new technologies and concepts are on the raise in the scenario of the industrial automation: Internet of Things (IoT), Industry 4.0, Industrial Internet of Things (IIoT) [67].

These IOT are closely related with AGVs, in most of the cases the AGVs need to receive and send information to a central controller related with: order assignations, traffic management, current state, etc. This information is mainly sent by proprietary protocols that normally spend more bandwidth that needed. In contrast to these ones, there exist modern protocols special designed to IOT communications like either MQTT or AMQP.

The MQTT (currently known as the ISO/IEC 20922 standard) has been originally designed for machine to machine telemetry data exchange in low bandwidth environments. MQTT supports publishing of messages to a broker on a topic. The message content is undisclosed to the broker, which uses the topic for message filtering and distributing to interested subscribers. Typically, the broker does not retain messages and store-and-forward mechanism are not explicitly considered. Three different data delivery models (and corresponding Quality of Service (QoS) levels) are defined for ensuring message reliability. When QoS Level 0 is used, messages are sent only once and the message arrival to their destinations is not checked. Therefore, messages can be lost. The QoS Level 1 enables a 2-way handshake mechanism, so that each message is sent at least once; however, if the message confirmation is lost, multiple deliveries can occur. Finally, QoS Level 2 sends the message exactly once utilizing a complete 4-way handshake. Obviously, in QoS Level 2, no message is lost but relatively longer end-to-end delays are expected.

AMQP was originally developed for the financial industry to furnish reliable exchange of business messages between applications. The underlying communication networks are supposed to offer high-availability and bandwidth and huge computational power is requested at the edge of the network, i.e. where field data are collected. AMQP is buffer-oriented, requiring high-performance servers and permits fragmentation of large messages. On the opposite, MQTT is stream-oriented, supporting low-memory clients and heterogeneous low-performance networks. Consequently, MQTT seems a more suitable choice for IIoT due to its efficiency; possible deficiencies can be solved at the application level [68]

4.1.3.3 Utility of big data for AGVs application

In the automotive sector, one production line stopped for one minute or one meter produces thousands of euros of cost. These figures indicate the need to increment the efforts to detect and prevent the problems in the production and logistic chains before their occurrence. Currently the traditional conveyors are being replaced by AGVs due to their flexibility, thus the prediction of the failures in the AGVs is a factor key in the productiveness of the Industry 4.0 factories. In this line, the datamining and big-data techniques in the predictive modality let us characterize predictive models by classification and regression tasks. By the classification tasks, it is assumed that data belong to different classes in function on their attributes and the target is to make classifiers to assign right label to unlabeled classes. The regression tasks are like the classification ones but in this case the attribute to be determined is not discrete but continuous.

In the other way, the bigdata techniques provide configurable performance indicators of the installations and the machines, this valuable information let us strengthen the procedures of feedback to design and continuous improvement.

4.1.4 Virtual factory simulation and planning

The virtual factory created by Visual Components solutions will mirror the real factory to create a digital twin that will allow to plan and validate new production alternatives to enhance productivity.

The concept of digital twin was first introduced by Grieves at one of his presentations about product lifecycle management in 2003 at University of Michigan. According to Grieves, digital twin contains three parts: physical product in real space, virtual product in virtual space and the connections of data and information, which ties the virtual and real product together (see Figure below). Virtual product is presented as a rich representation of a product that is virtually indistinguishable from its physical counterpart. [69]

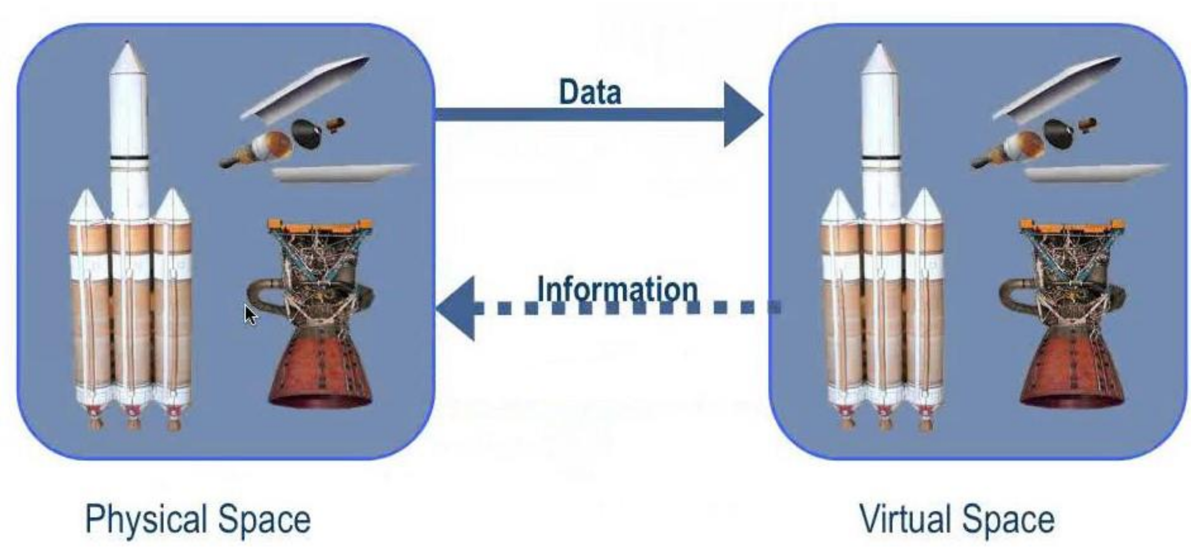


Figure 4-1 The concept of digital twin (modified from Grieves, 2014)

Virtual and physical model can be simultaneously viewed and compared providing several benefits, especially in the manufacturing point of view. Instead of simulating what should happen on the factory floor, digital twin replicates what has actually happened, at every step-in manufacturing process. This communication would happen real-time or near real-time, and it would transfer information, such as raw sensor data from physical object to virtual object. Information flow is not only one directional, but it can also be transferred from virtual object to physical object. This information can be, for example, data used for device control.

The digital twin capacity supports three powerful tools in the human knowledge kit. These tools are conceptualization, comparison and collaboration. Humans prefer to visually conceptualize situations, rather than look at a table of numbers, reports or other symbolic information. The digital twin enables information to be visually reviewed. Comparison is efficient tool of reviewing products. Physical product information can be visually and effectively compared to ideal characteristics of virtual product information. Collaboration provides more expertise, more variability of perspectives and improved problem-solving capabilities. Digital twin allows sharing ideas and conceptual designs, which can be easily distributed between the different shareholders to be visualized, analysed and improved. This allows global comparison between factories, which results in capabilities to improve manufacturing solutions immediately across the globe. [69]

4.1.5 Standards for Smart Production Planning & Management

4.1.5.1 Network data sources

Network traffic data, combined with network device information, allows the prediction of impacts on performance for the Industry 4.0 environment, caused by the ICT service disruption [70]. This requires the expansion of big data analytics and machine learning techniques to allow the use of the above-mentioned data sources, and in more detail of:

- Network flows: network packet-aggregation technology, grouped in flows (e.g. Netflow_v9, IPFIX [71], tstat [72]) in order to classify the nature of the traffic, their behaviour and priority.
- Network device and IT metrics in the factory communication network: extensive information from network devices (wireless access point, gateways, etc.), IT servers' status, and multiple usage and performance metrics (CPU, memory, I/O) of the IT system.

4.1.5.2 5G network technologies

5G networks are expected to support the needs of a hyper-connected society, asking for very high data rate access, requiring a wider coverage, and offering an increasing number of almost permanently connected devices. To enable the most demanding services, 5G standards are committed to increase performance over current mobile networks with figures as challenging as:

- 1000x mobile data volume per geographical area, around Tbps/km2
- 1000x the number of connected devices, over a million terminals/km2
- 5x improvement in end-to- end latency, targeting figures lower than 5ms

To provide high performance, ultra-low latency and high bandwidth, 5G RAN will rely on novel architectures and technologies (small cells, massive multiplexing, flexible resource sharing, etc.). On the other hand, to meet required flexibility and dynamicity, concepts such as network virtualization, multitenancy (slicing) or network programmability are perceived as key levers for 5G. In this way several key fora, like ITU, 3GPP, ONF, IEEE or IETF, are working on the standardization of the different components to have a complete solution for 2020. TID is making special efforts in standardization bodies and open-source communities, with active contributions in the SDN arena at IETF and ONF, especially relevant in Wireless Transport Information Models (TR-532), Transport API and Open Disaggregated Transport Networks.

In addition, governments and public authorities are also fostering 5G technologies. As an example, Europe has an important program from the 5G Infrastructure Public Private Partnership, which is under the Horizon2020 Framework Programme, targeted to help the development of these technologies by the European industrial and research ecosystem. Telefonica is working on several of these funded projects, like 5G-Transformer[5], 5G-MetroHaul[6], NECOS[7], 5G X-Haul[8], 5GEX[9], 5G-TANGO[10], or the recently initiated 5G-VINNI and 5G-EVE. TID is also a founder member of 5TONIC[11], an open innovation laboratory where industry and academia work together in research and innovation projects regarding 5G, including technology and new business innovation models with verticals

4.2 Trial present scenario

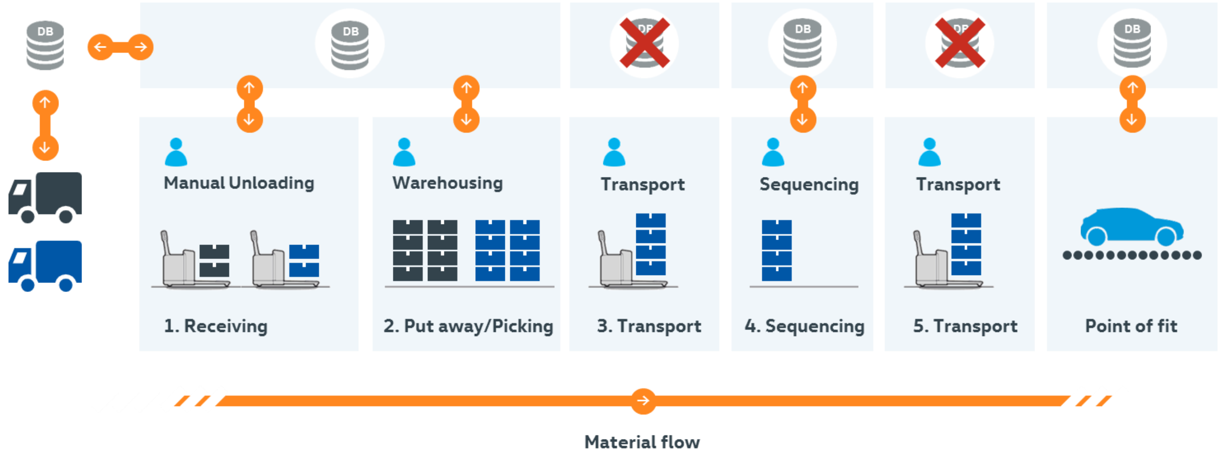


Figure 4-2 - System flow of current process

Currently the process is heavily reliable on manual processes and in addition to that, the operation is performed inside the factory, where space is limited.

On the receiving area trucks are traditionally unloaded by a manually forklift operation, and then the unit loads are transported to the warehouse where they will be stored either in shelve or block storage concept. System wise there is one database to control the parts coming from each truck and then a separate database, which registers the unloading, transportation and storing of the material in the warehouse.

An automatic line feeding system based on real vehicle demands generates parts call offs after interacting with real time stock data to replenish the points of use at commissioning areas called SUMA's, or directly at the assembly line using a pull methodology/concept. Deliveries are system supported until the point of fit.

[5] https://cordis.europa.eu/project/rcn/211067_en.html
[6] https://cordis.europa.eu/project/rcn/211077_en.html
[7] https://cordis.europa.eu/project/rcn/212036_en.html
[8] https://cordis.europa.eu/project/rcn/197339_en.html
[9] https://cordis.europa.eu/project/rcn/197346_en.html
[10] https://cordis.europa.eu/project/rcn/211063_en.html
[11] https://www.5tonic.org/

The next step is selecting the picking process for the correct sequencing in the SUMA. Here, the operator follows an electronic system for picking the parts according to the vehicle sequence on the production line. These operations are executed under the principles of the lean production system.

Conventionally the racks with the parts are moved from the picking/sequencing area to the point of fit through trolleys attached to tow tugs. This operation is exclusively performed by the operator and is achieved without resorting to any database, therefore any control or analysis of the process is not possible for the moment.

All SUMA line-feeding processes consist in establishing a circuit of deliveries of full unit loads and the collection of the empty trolleys and boxes used in the replenishment process for the assembly line.

Finally, the parts are manually delivery at the point of fit by the line-feeding operator.

4.3 Weaknesses and bottlenecks

When focusing on different areas t the following weaknesses/bottlenecks are highlighted:

Receiving / Manual unloading

1. Manual scan of each unit load (label)
2. Time used on cargo checking and visual incoming inspection done by each operator
3. All repetitive activities which have a direct impact on the operators' health and safety (consequence: medium to long term absenteeism)

Warehousing

1. Non-optimized warehouse arrangement due to physical constrains from the current operational machinery (optimization of the distance between each isle with AGV forklifts)
2. Although there are management systems to manage the warehouse it lacks a precise preventive feature to signal bottlenecks, a cognitive function to learn from historical data and an optimization tool to exhibit optimal storage scenarios.

Transport (to sequencing/delivery point)

1. The transport to delivery point is manual and it is not controlled
2. Data is available but is not fully integrated in the complete flow

Sequencing

1. The sequencing process is supported by systems that guide the operator throughout the operation. On the one hand, this step of the logistics process is sheltered from errors but on the other hand; this stage is overloaded with manual repetitive activities.

Transport (or point of fit)

1. The driverless vehicles (AGV's) are often limited to a fixed circuit therefore, dynamic routes do not exist.

Additional details are provided in the table below.

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA[1]	IMPACT IN THE COMPANY
Manual Operation	Manual unloading at Receiving implies the use of several operators that need to be aware of the type of material being delivered and how it should be handled and stored.	Logistics	This poses important challenges to the efficiency of the process in terms of time and accuracy as well as to the health of the operators especially when dealing with some heavy parts.
Data collection	Manual scan of each unit load is required at Receiving. During transport, no data is being collected.	Logistics	Manual scan contributes for loss of time in the process as well as potentiates the existence of errors. Impossible to check the location of a specific item during transport phase.
Data integration	Data is available at different phases (receiving, warehousing, sequencing and point-of-fit) but not fully integrated.	Logistics	Contributes to the possibility of inconsistencies and errors as well as it constitutes a barrier for the process optimization.
System communication	AGVs do not have communication capabilities.	Logistics	AGVs are limited to a fixed route therefore compromising process flexibility.

4.4 Trial future scenario

The envisaged future scenario aims at achieving a full integration of the material flow, from receiving up to the point of fit. Figure 4-3 shows the system flow integration as it is foreseen in VWAE.

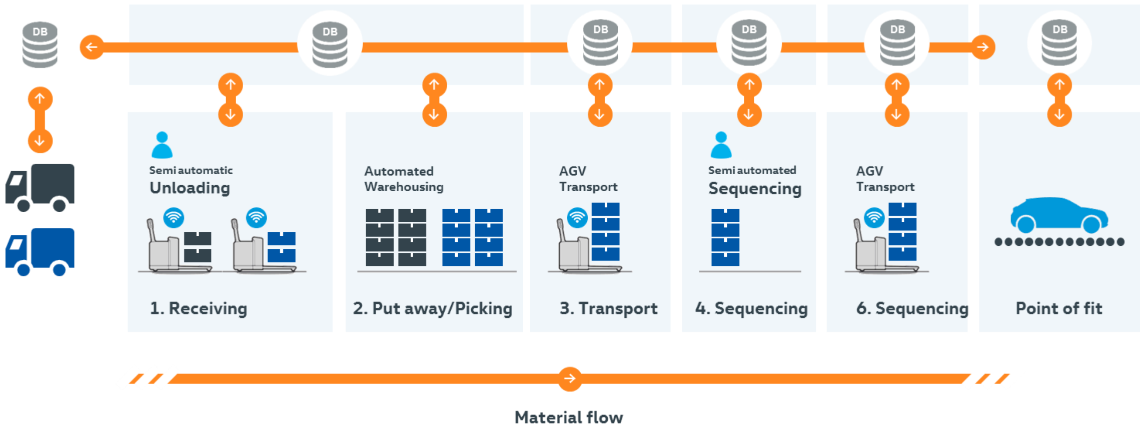


Figure 4-3 - System flow of integration 4.0

The main objective is to eliminate human intervention or at least reduce to a minimum at all phases from receiving up to the point of fit. The medium to long term vision is that activities will be fully, or at least to a large extent, automatized, namely:

- Automatic scan of each unit load (label)
- Automatic cargo checking and inspection
- AGVs with geolocation and communication capabilities will be used at all transport phases
- Picking Mobile Robots will be used for warehousing activities

The picking mobile robot will have the flexibility to adapt to different parts which will be linked and orchestrated by the existing sequencing system. The main objective is to lay down the pieces onto the rack in the correct order. The rack will be then picked up by a system composed by a set of AGV's which will handle the automated transport of these racks to the point of fit. Another aspect to consider is the monitoring functionality of the entire system available to the supervisor of the concept. This monitoring system will enable the control and management of all modular components as well as the regulation of key process indicators. Figure 4-4 shows the envisaged future scenario.

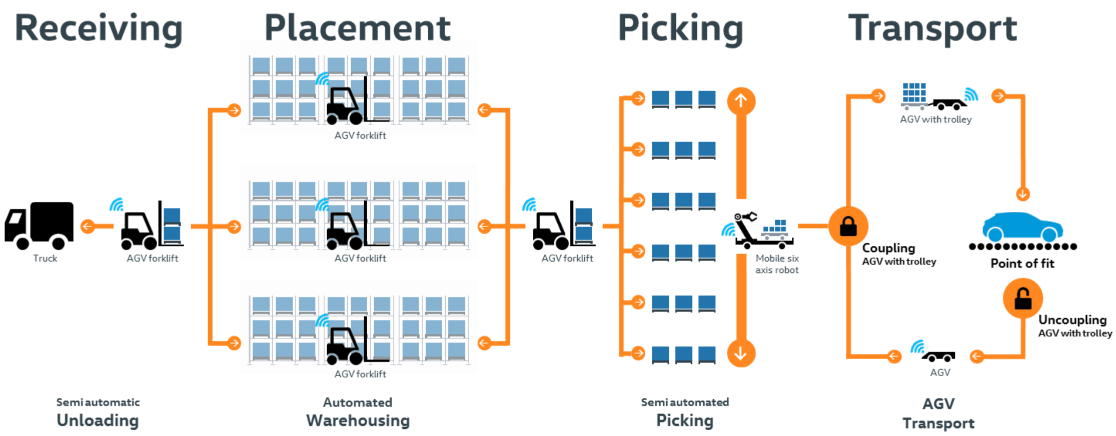


Figure 4-4 – Envisaged future scenario in detail

As the trial is quite ambitious from the point of view of focus processes and resources that would need to be available (robots, AGVs, etc.), the trial will be divided into several steps and each step will progress according to the results obtained at the previous steps.

Note that, for demonstration purposes, VWAE will focus on a specific part, the batteries, trying to model their complete flow. The goal of the internal logistics supply chain is to reliably deliver the part to the point of fit; however, there are parts where this can be achieved relatively easily and others where it is quite complex. Some parts due to physical complexity are quite cumbersome to handle and this is reflected in its process, therefore, parts with this characteristic may not be suitable to choose because they raise undesirable obstacles on the innovation forefront towards industry 4.0 concepts. The batteries then pose as a solid option to tackle since its shape is quite uniform making them easy to pack, store, handle and transport. Also, batteries cover the full scope of Volkswagen Autoeuropa production since they are installed in every car. For those reasons, it is believed that these parts should be considered as vessel for the dissemination of innovative processes since they are quite permeable to experimentation.

The first step will be to model the Receiving phase in order to obtain a sort of digital twin that will enable its simulation. Once the model is built, it will be capable of being simulated in order to: (i) test accuracy, (ii) test different scenarios (e.g. using different solutions for label scanning) and (iii) select the most appropriate approaches (e.g. the one that enables a decrease on the time spent on scanning).

The same approach will be applied sequentially to other process phases.

4.5 Expected results

The expected results for this trial can be divided into Technical Results (TR) and Business Results (BR). The first ones are closely related with the bottlenecks/weaknesses that were identified whereas the second ones derive from the general objectives identified by VWAE. The following tables identify these results (both technical and business) as well as proposes a set of Key Performance Indicators (KPIs) that will help assessing the success of the proposed approach.

Note that, for the Business Results, the expected outcomes considered are reflecting the focus on the logistics area. VWAE considers that, once similar strategies are applied in the complete process (including production), the impact will be higher.

TR nº	Result description	Bottleneck/ weakness	KPI
1	Detailed complete material flow for a specific part (Batteries)	Manual Operation	Simulation of semi-automatic mechanisms at unloading: at least 1
			Simulation of semi-automatic mechanisms at warehousing: at least 1
			Simulation of automatic mechanisms at transport: at least 1
2	Full data collection along the complete material flow for a specific part (Batteries)	Data collection	Automatic data collection mechanism at Receiving: at least 1
			Simulation of automatic data collection mechanism during transport: at least 1
3	Extraction of new/improved knowledge about the process	Data Collection	Recognition of new patterns (e.g. material quality): at least 1 new pattern
			Identification of new correlated events (e.g. problems with a specific supplier): at least 1 new event
			Definition of more efficient strategies to eliminate problems: at least 1 proved strategy
4	Complete specification of the system functionality	Data Integration/	Technical description of the functions performed: complete material flow of batteries

	including required interfaces	System communication	System Blue Print: all involved systems Description of interfaces: all required IT interfaces
5	Complete specification of the communication approach	System communication	Description of the necessary communication technologies and protocols between different systems: all involved systems
6	Detailed technical specification for the Data Management System	Data Integration	Description of the necessary data interoperability mechanisms: all related data

BR nº	Result Description	KPI
1	Increase of market share	Models produced within VWAE: +1 new model and or product update for current model
2	Improved Operational Efficiency	Number of stops/delays: - 10%
3	Improved Productivity/ Reduce overall production costs	Reduction of overall costs: - 1 - 5M€ (considering only logistics costs) Total cost of ownership (stock costs): - 10%
4	Reduce time to market	Time to market: - 15% (considering only logistics costs during launch phase)

4.6 Execution plan

The execution plan proposed for a first stage is to gather as much information as possible of the logistics internal supply chain at Volkswagen Auto Europa (namely layouts, warehouse and supermarket locations, etc.) this will allow a first shot at simulation although at a bare bones level, i.e.: rough location of main process points and rough definition of use case in the simulation. In parallel data is being collected daily from the different business processes in preparation for the data analytics stage of the project.

During this stage, visits were scheduled to Volkswagen Autoeuropa so that the whole group could have a better understanding of the process, as well as training sessions on Visual Components software. As of the date of this report, more training sessions are to be scheduled on Visual Components and also a workshop is planned at 5Tonic to explore the integration of AGV data and network fidelity onto Visual Components.

At a second stage, the efforts are to be focused on the fine tuning of the simulation, namely the hit/collision detection between the different components, replication of AGV routes, definition of possible routes for the transport process and the introduction of other components that have influence in the process.

This will also be the appropriate timing to introduce the actual data gathered from the different processes into the simulation; it will then show how faithful the system can be in comparison with reality.

The third and final stage should focus on the progressive work towards a robust digital twin were simulations of the internal supply chain of Volkswagen Autoeuropa (use case batteries) could come as close as possible to reality, enabling the study of different scenarios and their analyses between them.

There are also a handful of side tasks which are underway since the start of the project. They are related with market studies of suitable technologies which could help the level of automation and digitalization of the internal supply chain at Volkswagen Autoeuropa, data viability studies for algorithms and network solutions to improve communications between systems and line feeding equipment i.e.: AGV's.

5 Trial 4: +GF+

5.1 State of the art

5.1.1 State of the art on data analytics for predictive maintenance

Nowadays a lot of sensors have been developed from the concept of the predictive maintenance. For example, Watchdog AgentTM adopts smart prognostics algorithms to find when the observed process, or equipment, is going to fail or degrade to the point when its performance becomes unacceptable, and what the cause of the observed process or machinery degradation is[12].

In terms of model-driven approach, major methods used for predictive maintenance are system of differential application, rule-based expert systems, finite-state machines and qualitative reasoning. On the other hand, in data-driven approach, linear regression, Kalman filters, neural networks, decision trees and support vector machines (SVM) are widely used[13].

Even if there is a good number of machine learning algorithms today, neural networks, decision trees and Support Vector Machine (SVM) algorithms are usually used for the maintenance. Neural network is the algorithm usually using pattern recognition that exploits virtual neurons imitating humans to learn and simulate the mathematical model. Neural networks usually show good performance for pattern recognition, but it needs innumerable calculations and raw data. Decision tree algorithm is one of the useful algorithms for supervised learning. A model of the decision tree algorithm consists of making up interior nodes, which correspond to one of the input variables. The leaf is representing a value of the target variable given the values of the input variable by the path from the root to the leaf. SVM is a geometric algorithm to find a support vector to define the optimal hyperplane. This algorithm is usually used for writing recognition, object recognition on a picture and so on. In addition, in terms of representation of product states, a Bayesian network and Markov random field are useful in this study. A Bayesian network represents a set of random variables and their conditional dependencies via a Directed Acyclic Graph (DAG), and it can compute the probabilities of the presence of a failure. A Markov random field is similar to a Bayesian network in its representation of dependencies, however, it is undirected and cyclic, whereas a Bayesian network is directed and acyclic.

Especially, since the present situation of GFMS does not allow application of supervised-learning or semi-supervised learning approach and maintenance annotation is not available, this research should deal with the method of how to consider the new types of failure or fault which does not fit into the existing or predicted failures or faults. Therefore, this research requires the incremental

[12] Lee, J., Ni, J., Djurdjanovic, D., Qiu, H., & Liao, H. (2006) Intelligent prognostics tools and e-maintenance. *Computers in industry*, 57(6), pp. 476-489.

[13] Poongodai, A. & Reader, S. B. (2013) AI technique in diagnostics and prognostics. *International Journal of Computer Applications*.

classification. Incremental learning is a topic of major interest in machine learning defined as learning task which is incremental if the training examples used to solve become available over time, usually one at a time [14]. Various researches have proposed incremental learning algorithms. For instance, Fung and Managasarian (2002) addressed the incremental SVM [15] and Pang at al. (2005) proposed incremental linear discriminant analysis [16]. In addition, semantic technologies could be applied for incremental classification [17].

In particular, a detailed scan of main competitors and customers of GFMS has been done with the conclusion that most of them are implementing solutions or platforms which will allow gathering and monitoring data remotely from machines, but currently none of them is showing evidence of a predictive maintenance application as described in this proposal. Examples are DMG, with the Celios platform, using deterministic models for spindles maintenance, Mazak, marketing a Smartbox for connectivity applications but with no similar analytics beyond monitoring and production statistics. Main competitors in EDM, like Makino and Sodick, announce remote connectivity products but no predictive maintenance applications as aimed by this project.

According to ABI Research forecasts, the market size for maintenance analytics will reach $24.7 billion in 2019, driven largely by adoption of predictive analytics and M2M connectivity[18]. In this context, as the potential in this market is growing year-by year, there are certain attempts of large multi-national solution provider companies such as SAP and IBM along with collaborative centers addressing the issue.

On that regard, the "Center for Intelligent Maintenance Systems" envisions the future of maintenance as a system that enables equipment to achieve and sustain near-zero breakdown performance with self-maintenance capabilities (self-aware, self-predict, self-compare, and self-configure), and ultimately to realize the autonomous transformation of raw data to useful information for improved reliability, productivity, and asset utilization[19].

Besides, predictive maintenance software solutions from IBM access multiple data sources in real time to predict asset failure or quality issues to avoid costly downtime and reduce maintenance costs. Driven by predictive analytics, these solutions detect even minor anomalies and failure patterns to

[14] Giraud-Carrier, C. (2000). A note on the utility of incremental learning. *Ai Communications*, 13(4), 215-223.

[15] Fung, G., & Mangasarian, O. L. (2002). Incremental support vector machine classification. In Proceedings of the *2002 SIAM International Conference on Data Mining*. pp. 247-260

[16] Pang, S., Ozawa, S., & Kasabov, N. (2005). Incremental linear discriminant analysis for classification of data streams. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics),* 35(5), pp. 905-914.

[17] Grau, B. C., Halaschek-Wiener, C., Kazakov, Y., & Suntisrivaraporn, B. (2010). Incremental classification of description logics ontologies. *Journal of Automated Reasoning*, 44(4), pp. 337-369.

[18] https://www.abiresearch.com/press/maintenance-analytics-to-generate-247-billion-in-2/

[19] http://imscenter.net/

determine the assets and operational processes that are at the greatest risk of problems or failure. This early identification of potential concerns helps deploy limited resources more cost effectively, maximize equipment uptime and enhance quality and supply chain processes, ultimately improving customer satisfaction. However, following a top-down approach for its analysis, it provides generic solutions on an aggregated level rather than the breakdown and compensation mechanisms for a specific machine tool[20].

Last, "SAP Predictive Maintenance and Service, cloud edition" maps the data from the devices to the master data of the asset in the respective business systems, and then monitors the incoming data and generates alarms once critical thresholds are exceeded. Those alarms then triggers follow-up actions like service or maintenance orders in business systems to ensure smooth integration with established business processes[21]. There is however no evidence for applications in the machine tool domain and their offer remain far from real context of the technologies provided by GFMS.

5.1.2 State of the art on ontology engineering

The ontology engineering is the general term of methodologies and methods for building ontologies. Ontology engineering refers to "The set of activities that concern the ontology development and the ontology lifecycle, the methods and methodologies for building ontologies and the tool suites and languages that support them" (Suárez-Figueroa et al., 2011). The results of ontology engineering provide domain knowledge representation to be reused efficiently, and prevent waste of time and money which are usually caused by non-shared knowledge. In addition, it helps Information Technology (IT) to operate with interoperability and standardization.

✓ *Semantic Modelling*

Ontology represents the nature of being, becoming, existence, and so on in the way of philosophy. Among various definitions of ontology, the most well-known definition of ontology is : "ontology is an explicit, formal specification of a shared conceptualization of a domain of interest" (Gruber, 1993). In other words, ontology is the machine understandable meta-model which defines different kinds of concepts and their relations based on the consensus knowledge among not only the members of the domain but also computers.

Ontology represents the following ideas together[22]:

- Semantic modelling can help defining the data and the relationships between entities.
- An information model provides the ability to abstract different kind of data and provides an understanding of how the data elements are related.

[20] https://www-01.ibm.com/software/analytics/solutions/operational-analytics/predictive-maintenance/

[21] http://scn.sap.com/community/internet-of-things/blog/2014/11/11/launch-of-sap-predictive-maintenance-and-service

[22] Calero, C., Ruiz, F., & Piattini, M. (Eds.). (2006). Ontologies for software engineering and software technology. Springer Science & Business Media.

.

- A semantic model is a type of information model that supports the modelling of entities and their relationships.
- The total set of entities in our semantic model comprises the taxonomy of classes that is used in our model to represent the real world.

The main objective of semantic modelling techniques is to define the meaning of data within the context of its correlation, and to model the real world in the abstract level. The benefits of exploiting semantic data models for business applications are mainly as follows:

- Avoiding misunderstanding: by providing a clear, accessible, agreed set of terms, relations as a trusted source and discussions, misunderstandings can easily be resolved.
- Conduct reasoning: by being machine understandable and through the usage of logic statements (rules), ontologies enable automatic reasoning and inference which leads to automatic generation of new and implicit knowledge.
- Leverage resources: by extending and relating an application ontology to external ontological resources, via manual or automatic mapping and merging processes, the need for repetition of entire design process for every application domain is eliminated.
- Improve interoperability: semantic models can serve as a basis for schema matching to support systems' interoperability in close environments where systems, tools and data sources have no common recognition of data type and relationships.

Ontologies provide formal models of domain knowledge exploited in different ways. Therefore, it is important that ontology plays a significant role for many knowledge-intensive applications.

Depending on corresponding languages, different knowledge representation formalisms exist. However, they consist of the following minimal set of components and share them:

- **Classes** represent concepts, which are taken in a broad sense. For instance, in the Product Lifecycle domain, concepts are: Life Cycle phase, Product, Activity, Resources, Even, and so on. Classes in ontology are usually organized in taxonomies through which inheritance mechanisms can be applied.
- **Relations** represent a type of association between concepts of the domain. They are formally defined as any subset of a product of n sets, that is: $R \subset C1 \times C2 \times \ldots \times Cn$. Ontologies usually contain binary relations. The first argument is known as the domain of the relation, and the second argument is the range.
- **Formal axioms** serve to model sentences that are always true. They are normally used to represent knowledge that cannot be formally defined by the other components. In addition, formal axioms are used to verify the consistency of the ontology itself or the consistency of the knowledge stored in a knowledge base. Formal axioms are very useful to infer new knowledge.

For instance, Energy Efficiency at Buildings domain could be that it is not possible to build a public building without a fire door (based on legal issues).

- **Instances** are used to represent elements or individuals in an ontology.

As a Design Rationale (DR), ontology can be used as follows[23]:

Level 1: Used as a common vocabulary for communication among distributed agents.

Level 2: Used as a conceptual schema of a relational database. Structural information of concepts and relations among them is used. Conceptualization in a database is nothing other than conceptual schema. Data retrieval from a database is easily done when there is an agreement on its conceptual schema.

Level 3: Used as the backbone information for a user of a certain knowledge base. Levels higher than this plays roles of the ontology, which has something to do with "content".

Level 4: Used for answering competence questions.

Level 5: Standardization 5.1 Standardization of terminology (at the same level of Level 1) 5.2 Standardization of meaning of concepts 5.3 Standardization of components of target objects (domain ontology). 5.4 Standardization of components of tasks (task ontology)

Level 6: Used for transformation of databases considering the differences of the meaning of conceptual schema. This requires not only the structural transformation but also semantic transformation.

Level 7: Used for reusing knowledge of a knowledge base using DR information.

Level 8: Used for reorganizing a knowledge base based on DR information.

5.1.3 INENDI - Inspector

INENDI is described in section ***¡Error! No se encuentra el origen de la referencia. ¡Error! No se encuentra el origen de la referencia.*** of the VW trial.

[23] Mizoguchi, R. & Ikeda, M. (1998) Towards ontology engineering. *Journal-Japanese Society for Artificial Intelligence*, 13, pp. 9-10.

.

5.1.4 SCILAB Opensource & SCILAB Cloud

. SCILAB is described in section ***¡Error! No se encuentra el origen de la referencia. ¡Error! No se encuentra el origen de la referencia.*** of the VW trial.

5.2 Trial present scenario

The following flow diagram depicts in a simplified way the manufacturing process for the specific target of the GF pilot, that is the manufacturing of the milling spindles.

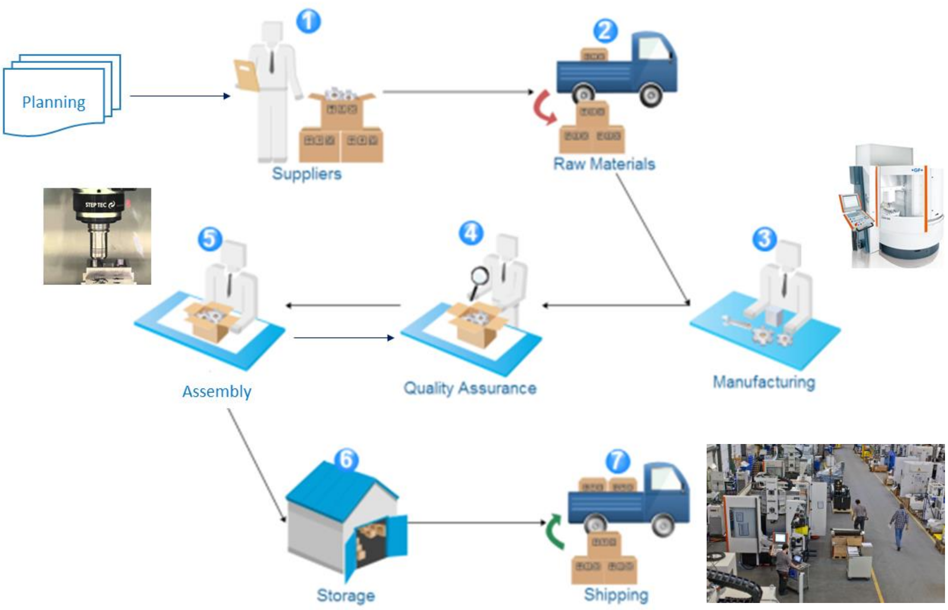


Figure 5-1 Manufacturing process of the milling spindles

Following a planning stage, supplies are scheduled and delivered to the spindle manufacturing plant, where critical elements are machined using GF Milling machines (stage 3). These critical parts are then controlled using measurement devices before going into the assembly stage. Another quality assurance stage takes place after the assembly (stage 4-5 iteration) before the new spindle integrates the manufacturing flow of a new machine.

The current flow does not allow a real time update of the overall planning as the communication of data from the line is mostly paper-based. Additionally, there are several manual operations with no real time follow up and key information about the quality and speed of the process is lost. Basically, there is no easy access of data from every stage for a reliable control of the production and reduce the number of defected parts which are only detected at the quality control points.

5.3 Weaknesses and bottlenecks

This pilot has concerns as follows; i) the acquisition and storage of data, ii) data analytics on operating data and monitoring, iii) continuous evaluation and prediction of the health status of the equipment as cyberphysical system (i.e. transform descriptors extracted previously in relevant behaviour models, allowing to represent the ways of functioning of the machine and the evolution of the equipment condition over time for detection and prognosis of failures), iv) decision-making support by considering the context of use of the equipment.

To overcome these concerns, the main approach for GFMS pilot is based on the use of advanced enabling technologies for monitoring complex equipment. It requires integration capacity of heterogeneous data sources. In addition, its innovation lies on the incorporation of various domains of knowledge. ontology facilitates a multidisciplinary approach through representation of domain knowledge with its concrete definition and provides semantic interoperability. To recognize semantic context of data brings values as follows: i) End users can easily recognize and identify the meaning of individual data, and search meaningful data, ii) Software developers can harmonize data from various sources and request required data for each system components, and facilitate design of machine understandable entities on an intelligent engine, and iii) Requirements of end users can be satisfied since knowledge representation acts as a bridge between end users of a software platform and platform developers for design to meet requirements of end-users.

On the other hand, data does not have annotations describing machine status or maintenance history, even if an amount of data is available. In other words, the company has limited options to analyse manufacturing data, even if advanced machine learning techniques support finding the criteria or failure symptoms and optimizing the maintenance schedule. Accordingly, this research should deal with the method of how to consider the new types of failure or fault which does not fit into the existing or predicted failures or faults.Regarding the analytics side of the current scenario, these are few identified weaknesses:

- To be efficient, visual data analytics tools, such like INENDI Inspector, required to gather **representative enough amount of data** in order to clearly identify possible behaviors (normal / abnormal), this is all the more important when working with non-labeled data (unsupervised), which seems to be partly the case in this project.
- Traceability, which is a key point in that project, requires the ability to localize a root cause of a failure (defected part, process) in the production flow presented in the current scenario. Since this global flow is composed of several sub processes (7 stages, as described in 2.5), themselves feeding different databases, these databases have to be **filled with key (pivot) parameters**.
- **Missing information/parameters** which makes that no clear pattern could be identified in our data visual inspection:

 - Measurements devices used for controlling the workpiece (Coordinate Measuring Machines) checks few critical points, some important information on the machined geometry may be lost.
 - Milling is a complex process which involves mechanical / chemical / thermal reactions. Furthermore, workpieces are submitted to initial constraints before the milling (residual stresses coming from the history of the material).
- **Difficulties to identify patterns linked to abnormal behavior (defects)** in time series coming from milling process (cutting forces, feed rate, displacement/velocity/acceleration, coordinates …).

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA
Heterogeneous data sources	Constrains for integration	Manufacturing
No maintenance annotation	Limited options of data analytics	Manufacturing
Lack of data	No history recorded	
Missing binding key	Missing pivot parameter to navigate between databases	
No relevant parameters	Missing relevant characteristics	
No pattern clearly identified	Difficulties to extract pattern for defining predictive models	

5.4 Trial future scenario

The following figure represents the big picture of the future scenario tackling all the weaknesses and exploiting the opportunities provided by Boost 4.0 enablers for a zero defect factory for milling spindles for GF machines.

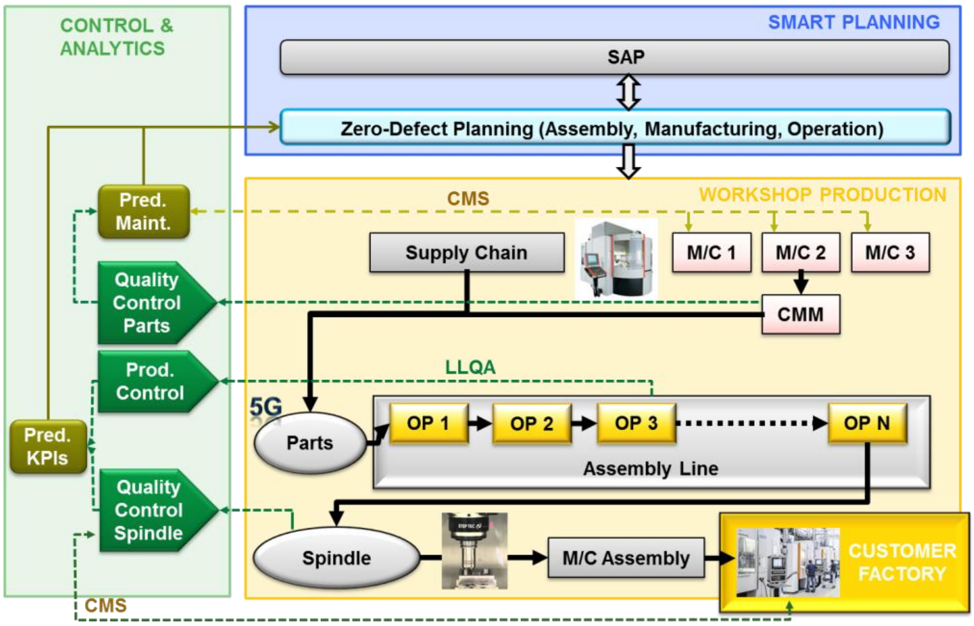


Figure 5-2 Big picture of the future scenario

The different elements of the future pilot integrate seamlessly data from manufacturing stages for the milling spindles, from the predictive maintenance and process monitoring of machines used for the manufacturing of components so to improve the efficiency and the quality of the parts, to the real time traceability of components in the assembly stage, so to provide a framework for the development of a smart, adaptive planning system leading to a zero defect, high productivity manufacturing process.

The following elements integrate the full concept:

Smart Planning

- Insure quality specifications
- Updated by predictive maintenance and prediction system of quality KPIs
- Adapts manufacturing parameters based on KPIs to insure the quality and productivity specifications

Production process

- Manufacturing of necessary spindle components in a first stage
  - Components with very high accuracy demands
  - Inspection of produced parts on CMM machines
- Additional components provided by suppliers in second stage
- Several workstations for the assembly of the spindle

Cognitive zero-defect framework

- Quality control of produced parts
    - Geometric quality inspection using CMM
    - Surface integrity measurements
- Quality control of assembled spindle
    - Tests to confirm the correct functionality of the spindle
- Production control
    - Manufacturing process transparency through localization of parts, digitalised and interactive manuals, operation plans update and quality specifications measurements to increase efficiency and process traceability
- KPIs prediction
    - OEE, Throughput, Supplier's Quality Incoming, Quality tracking
- Predictive maintenance – sensor and part data based - to ensure the availability and quality of the equipment, over entire life-cycle
- Cognitive framework implemented in the Smart Planning module delivering execution updates for zero defect manufacturing

Zero defect manufacturing of spindle components

The ambition of the GF pilot is to develop an adaptive machining system in automated cells, integrated into an open cloud and data analytics infrastructure, taking into account actual information of machine condition and performance in accuracy, surface quality, productivity and sustainability. The eco-systems will be able to diagnose in real time the state and performance of the machine and correct deviations through updated simulation models and planning systems.

The system should be automatically configurable and secure, in order to allow the integration of data coming from different hardware and not bound to one single supplier but allowing an eco-system, seamless integration of all relevant data in a specific common data and analytics space allowing appropriate data access authorizations.

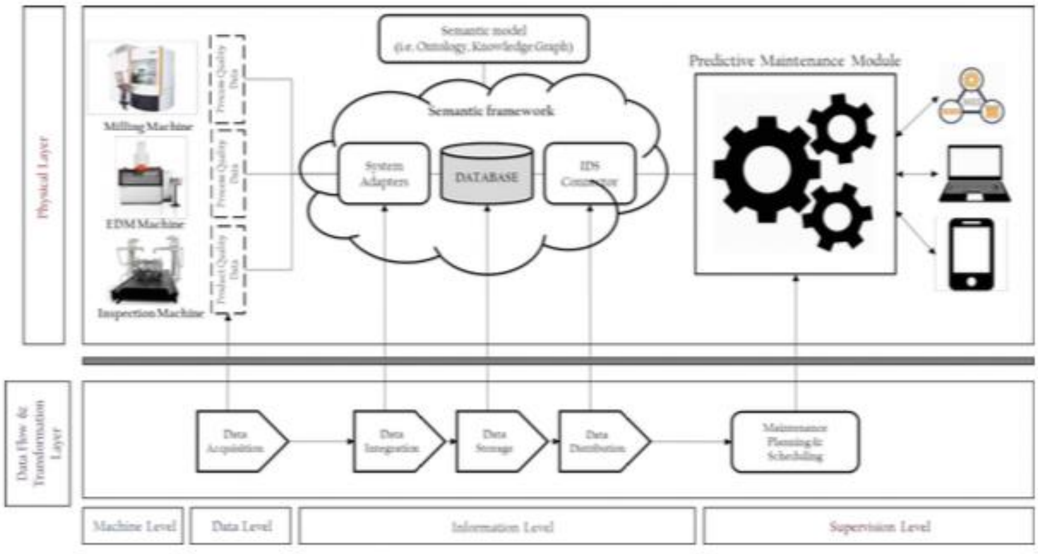


Figure 5-3 Adaptative machining system

The data analytics infrastructure will be designed as shown in the following figure. The continuous improvement loop (data / analytics / computation) will be performed in order to provide information for the workshop production monitoring that will help the decision making.

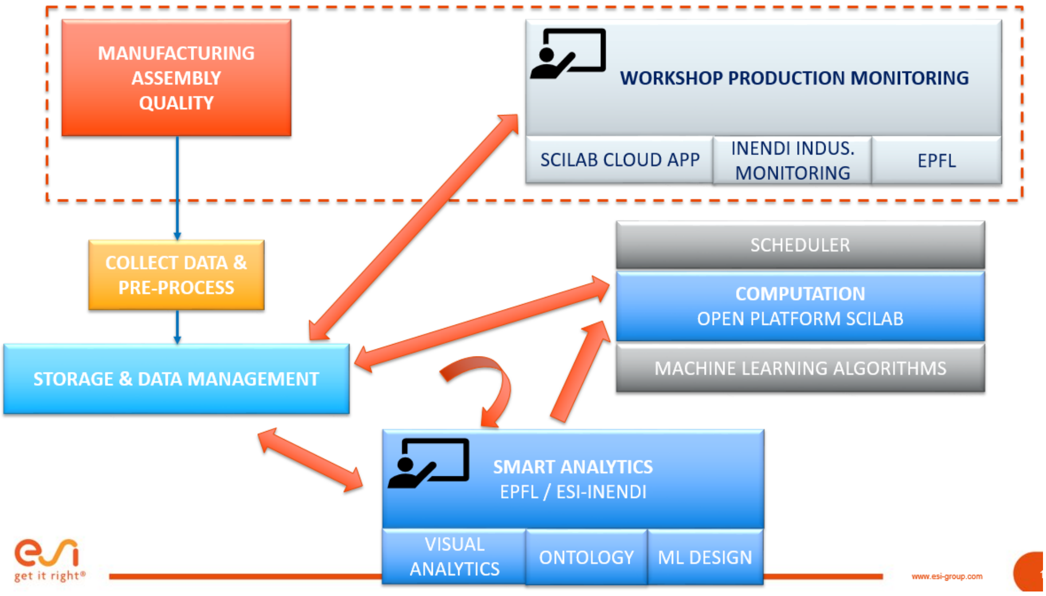
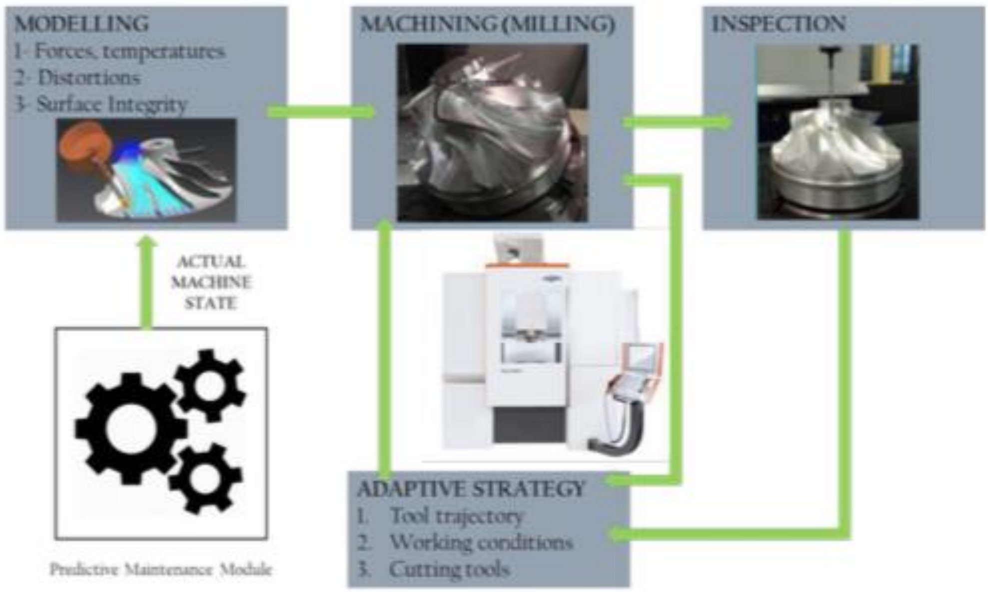


Figure 5-4 Data analytics infrastructure design

This coupled predictive-quality maintenance system will be able to provide more information than a standard predictive system based solely on machine sensor data, and will deliver correlations between quality/accuracy issues and potential component failure root causes. Such correlation will help to:

- Update maintenance planning systems by predicting critical machinery states

- Update simulation models for adapting the machining process to the new equipment condition before going above the critical threshold



For mould and die applications the critical stage is the milling operation of the electrodes to be used for mould machining with EDM technology. Updating information of both the real geometry of the electrode and the strategies for subsequent operations will eliminate error propagation and accuracy degradation over time.

5.5 Execution plan

Early Phase: Data Analysis & Monitoring (Month 1 to 16)

Sensor and Data management systems setting (M1 – M10)

- Set Sensors, exchange protocols and data management system to ensure data collection
    1. Workshop machines: GF, DMG, Studer and CMM machines
    2. Assembly chain

LLQA 4.0 development (M8 – M12)

- Design Thinking of LLQA upgrade towards interactive data aggregation tool from Assembly chain.

Database and data collection for already existing processes (M10 – 12)

- Create a database by running trial machines during months
- Identification of relevant parameters on available systems

Data analysis & Machine Learning algorithms for predictive maintenance & quality (M10 – M14)

- With ESI-INENDI (Inspector) and EPFL collaboration
    - Data analysis
    - Design predictive maintenance models
    - Define monitoring apps content

Machine Learning algorithms training and monitoring dashboard (M12 – M16)

- Based on created smart data basis by ESI-INENDI, EPFL will:
    - Set up machining learning training
    - Define effective algorithms for specifically developed dashboard monitoring

Production Phase: Production Workshop Optimization - Pilot (Month 16 to 24)

Data & KPIs monitoring analysis (M1 6– M20)

- ESI-SCILAB/INENDI/EPFL development of dashboard application to monitor and analyse KPIs of the whole production workshop

Control on assembly line (M16 – M22)

- Development of a Control strategy on the assembly line to ensure speed and quality of the process (GF/EPFL/ESI-SCILAB)

Planning business applications (M18 – 22)

- Development of a planning application based on Xcos module to plan activity all over the production workshop (from supply chain to final product) (GF/ESI-SCILAB)

Using KPIs and predictive maintenance output for smart process planning (M20 – M24)

- Leverage assembly line control and predictive maintenance model to ensure an efficient production workshop. (GF/EPFL/ESI-SCILAB)
- Goal to achieve is Zero-Defect Factory.

6 Trial 5: FIAT autonomous assembly line factory 4.0

6.1 State of the art

6.1.1 MindSphere

MindSphere is the open IoT Operating System provided by Siemens. The following schema shows the 3 layers architecture:

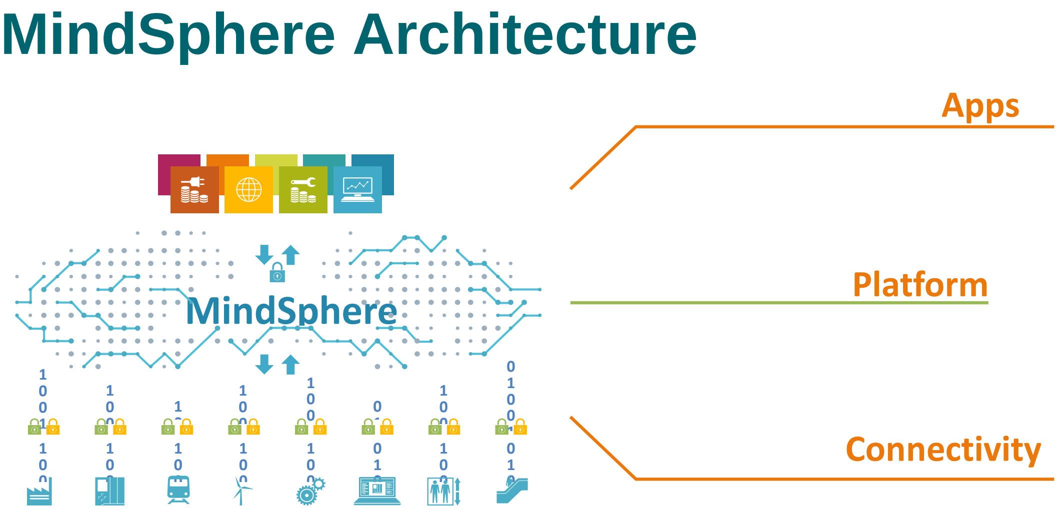


Figure 6-1 MindSphere architecture

For Connectivity, MindSphere provides multiple, varied and easy-to-implement connectivity solutions (both, hardware and software based) to be able to onboard a wide range of assets (Siemens and 3rd-party) in both brown-and greenfield environments.
MindSphere is a platform (PaaS) actually running on AWS' infrastructure and offer several APIs such as MindConnect API, Identity and Access Management, Asset Management, Event Management, Notification Service, Anomaly Detection, KPI Calculation etc.
Once developed a MindApp it is possible to push it into the Market Place and make it available for any MindSphere User or to keep it private for a specific tenant.

6.1.2 Prima

Figure 6-2 The laser Next

Laser Next is a 3D laser system designed, developed, manufactured and tested for the production of automotive components, particularly HSS parts. With the new 2130 model Laser Next is suitable also for large sized parts (e.g. new door ring concept).

First-class performance to grant lowest cycle times and excellent quality.

Higher OEE thanks to reduced downtime and maintenance.

Space saver, especially for multi-machine configuration.

Footprint:

- Less energy, less waste of material, no laser gases – Less CO2
- Higher laser wall-plug efficiency, less cooling capacity required, less heat generation – Less CO2
- Compact and automated – less factory space and logistics
- High efficiency – more output during machine run

Technical data:

<u>Work area</u>

X 3,050 mm Y 1,530 mm / 2100 mm Z 612 mm

A 360° B ±135° C ±12 mm

<u>Axis speed</u>

X, Y, Z 120 m/min (trajectory 208 m/min)

A, B 1.5 rev/s

<u>Laser source</u>

Fiber 3,000 W – 4,000 W

6.2 Trial present scenario

The industrial experiment will be a shopfloor equipped autonomous systems (AGVs) linking cells (PRIMA, FCA). The concept of autonomous production, where the traditional linear process is removed and Mobile Robots such as Automated Guided Vehicles (AGVs) and Collaborative robots with vision capabilities among others are included. But currently, these mobile robots have duties related to logistics (e.g. replenishment, preparation of components, etc) or manufacturing (e.g. carrying work in progress) and the control of fleets of such AGVs and their availability and reliability to respect cycle time and leadtime is crucial to ensure the stability and throughput of the production systems. However, planning and control of the mobile robots and monitoring and maintenance of the mobile robots are required due to currently there no specific approach to store and analyse data related to the missions of the vehicles, their wear-out and availability, taking into account the leadtime for delivery and the uncertainty related to the interaction with the presence of human operators.

One of the main objectives of the trial is to ensure that the new technology will be robust enough to avoid business interruption (e.g. stock-out, unwanted waiting or idle time for the machine), delays and reduction of throughput to transfer the autonomous production to the rest of the plants of FCA.

3D laser systems at the current state are designed, developed, manufactured and tested for the production of automotive components, particularly HSS parts.

The clamping system is fixed on the rotary table and each laser machine could just work a unique part number of the car because it is not possible to manage different clamping system on the same machine without changing the rotary table. Furthermore, the clamping system is mounted by operator, and every time that the part changes, the operator has to replace the clamping system for that specific new part: it means there must be at least one operator on a machine. This leads to an extension in loading and unloading time. The amount of data to be managed is very huge, about 80 kB every 300 msec: it means approximately **23 Gb** of data to analyse in one day.

DATA SOURCES: Stream –Yes, **Variety** – High, **Veracity** - Medium

Most of the time, the extrapolated data are not understood, which are therefore wasted.

Another weakness is the quality control: the quality check is done on a sample piece every 100 pieces. This implies that if the sample is compromised, will be considered as such as also the other 100 pieces. This is obviously a waste of time and materials that must be avoided.

6.3 Weaknesses and bottlenecks

The most common difficulties are often on the field level: heterogeneous devices, each one with specific communication protocols and data point list. Also, the security policy of the industrial field may represent a bottleneck for a procedural point of view.

Due to the confidentiality nature of algorithms (provided by FCA) and data (collected from FCA devices) the project may arise extra privacy topics that may impact also on the distribution of the application.

Industrial dark data: Understanding, organizing and using expansive data sets in new and better ways; lack of data continuity, uniformity, and standardization across product development lifecycle brings incredible challenges for data exploitation.

Speed mismatch between production process and information flow: One of the main challenges will be the reduction of time between the production process and information flow that has to be approximately equal to zero.

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA[1]	IMPACT IN THE COMPANY

			At processes (manufacturing) level and business level
Field heterogeneity	Complex connectivity	Manufactory	Deep field analysis (technical)
Security policy	Slow project development	Management	Deep network and policy analysis
Algorithms and data confidentiality	Confidentiality	Marketing and Sales	Additional agreement on marketing and sales actions

6.4 Trial future scenario

The following schema shows the evolution of the architecture that wants to be implemented:

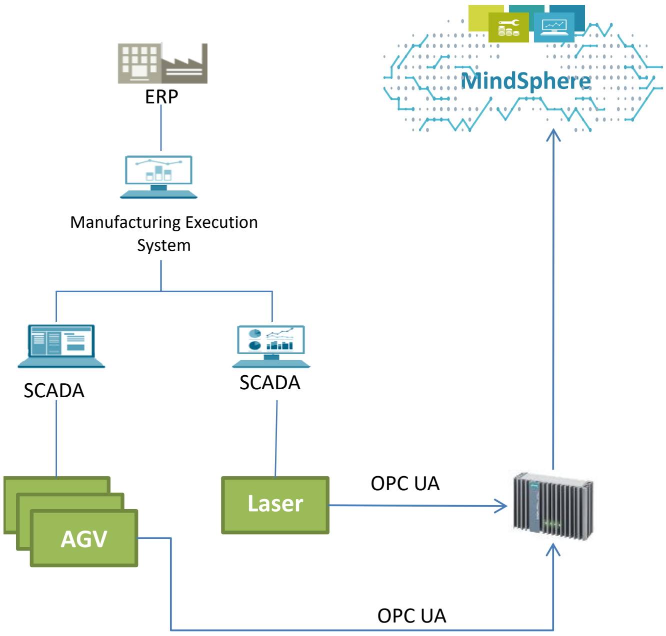


Figure 6-3 Evolution of the architecture

The idea is not changing the consolidate approach, so data will flow from the field to the SCADA and MES and ERP as per automation best practices; the innovation is to get out data from the field and send it directly to MindSphere, the cloud platform.

The scope of the project is to standardize the connectivity therefore the OPC UA protocol will be used as per OT best practice. Then a gateway (hardware or software) will be used to send secured data to

MindSphere. On top of the cloud platform a new application will be developed that transforms data in value (i.e. predictive maintenance, reporting analytics etc.)

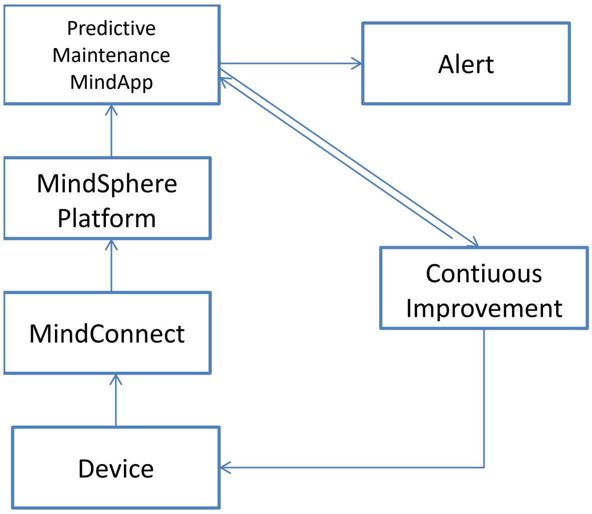


Figure 6-4 Reference architecture

The idea is to have:

- inter connected laser machine;
- vision system;
- AGV for pallet: clamping system is integrated on the pallet.

This way the vision system, AGV and laser machine are interconnected. Pallets will be brought on the working area of the laser machine. The vision system will recognize the part numbers of specific piece mounted on the pallets and it will communicate to the machine the right part program, and the position of the pallet inside the working area. After that machine will cut the part and the vision system will measure tolerances of the worked part. All data will be available on the cloud (MQTT Protocol). The machine, the vision system and the AGV will be connected through IoT of the factory or they will be managed through cyber physical systems.

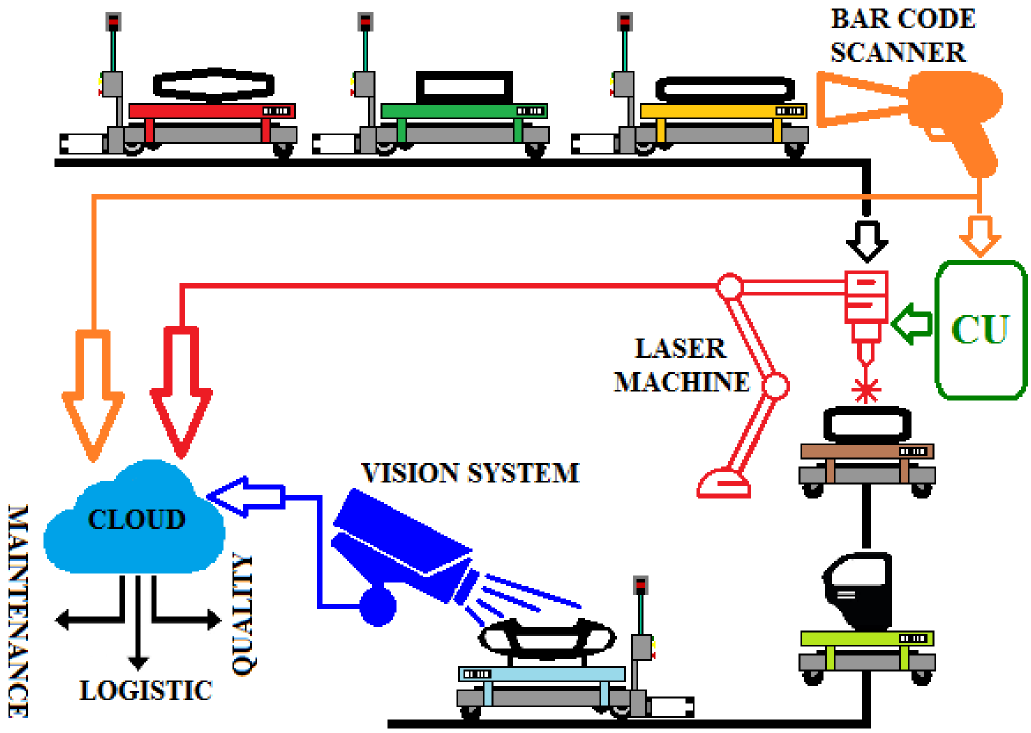


Figure 6-5 All data will be collected on the Cloud and available for all

6.5 Expected results

From a connectivity point of view, the standard technology and the best practice for data collection, data transfer and data storage will be defined.

From a functional point, of view the aim of the trial is to define procedure and best practice and best procedure for predictive maintenance and also to define a closed loop innovation process.

The trial will prove the maximum flexibility to potential changes in the demand or to issues/delays/changes in the logistics or productive systems by means of using available and new datasets (such as flows of components in the plants and their precise localization) ensuring business continuity, at the same time the over-dimensioned fleet of robots is reduced and the (big-)data are shared among the whole value chain (providers, maintenance services, etc.)

6.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 1: Smart Operations & Digital Workplace					

BUSINESS PROCESS 1 - Flexible and Smart Plant Operations through AGVs					
1	PoC for data acquisition, storing and visualisation				
2	Validation of the interaction between AGV and cell				
3	Trial of the interaction between AGV and cell				
4	Evaluation and KPIs collection				
BUSINESS PROCESS 2 - After sales service & Downtime reduction					
1	PoC for data acquisition, storing and visualisation				
2	Development of visual monitoring tools for AGV fleets and machines				
3	Trial of the visual monitoring tools				
4	Evaluation and KPIs collection,				
BUSINESS PROCESS 3 - Analytics & Predictive maintenance					
1	PoC for data acquisition, storing and visualisation				
2	Development of analytics for fleet status and quality level				
3	Development of interfaces and visual output of predictive tool				
4	Trial of the predictive maintenance tools				
5	Evaluation and KPIs collection,				
BUSINESS PROCESS 4 - Monitoring & Diagnosis					
1	Development of the IDS architecture for data exchange between data provider and data consumer				

2	Integration of IDS-based shop floor data cloud transfer				
4	Trial of the IDS-based shop floor data cloud transfer				
5	Evaluation				

7 Trial 6: PHILIPS Autonomous short-batch injection moulding production process

7.1 State of the art

7.1.1 Tengu

The Tengu platform[24] is a fully automated self-servicing PaaS for IoT and big data projects. Clients can use the Tengu platform to deploy a wide variety of big data technologies (moreover in the next subsection) and stitch them together, thus creating the complete backend necessary to run any big data project. Deployment can go towards any popular public cloud provider (Google, Amazon and Azure), private clouds such as VMWare and OpenStack, or even bare metal servers. Tengu abstracts everything necessary to communicate with these cloud providers wrt. Virtual machines, private networks and firewall rules to route access within the setup and from the outside to the components inside the deployment. Not only deployment is fully automated by the Tengu platform, but also installation of the necessary services, configuration of these components, and what makes the Tengu platform exceptional, the integration of all these technologies. This allows our clients to immediately get started with defining and developing the actual data flows and application components that drive the big data project. Furthermore, tools to work directly on the platforms data are easily integrated, which means data scientists can also immediately start experimenting with their algorithms, driving innovation and business insights light-years forward. Once the deployment is completely set up, Tengu also monitors and maintains all components, and provide high-level management tasks, e.g. auto scaling.

The Tengu platform is commercially available through the IMEC and IDLab's spinoff Tengu (http://www.tengu.io). It is currently used in many Belgian and European projects to manage either the backend or even to control frontend applications.

7.1.1.1 Integrated Big Data Technologies and other clustering technologies:

This section covers the different big data related technologies that are supported by the Tengu Platform. As one may notice all technologies are Open Source technologies.

- **Data Ingest and Message Systems**

  Integration with IDLab's Dyamand as generic low-level abstraction layer (moreover in a following subsection); Apache Kafka, at this moment the most important distributable Message

[24] http://www.tengu.io

Framework; Kafka HTTP REST interface; RabbitMQ and ActiveMQ thus supporting AMQP; Thingsboard supporting MQTT, Device Management, CoApp

- Data Analysis
  Support sthe full Hadoop Stack available by Apache Bigtop, most predominantly Apache Spark, Flink and Apex; Zeppelin notebooks and Jupyter notebooks to easily integrate with Spark, Flink and numerous data stores; Kafka Connect; Heron and Node-RED for lightweight data flows
- Data Storage and Distributed FS
  Most popular data store technologies are supported, RDBMS: MySQL, Postgress and MariaDB; NoSQL: MongoDB, Cassandra, ElasticSearch, ArangoDB, RethinkDB, InfluxDB, Prometheus and OpenTSDB. Next to these highly scalable data stores, we also support HDFS (with Yarn), GlusterFS, Ceph and Minio.
- Integrated Cluster Technologies
  Supports ArrangoDB Fox, a lightweight microservices framework. Next to this we can deploy a full Kubernetes cluster, the de facto standard in containerized scheduling and management. Addtionally, the deployment of dockerized containers on Kubernetes over Helm or via the Kubernetes API is supported.

7.1.2 Dyamand

Dynamic, Adaptive Management of Networks and Devices (DYAMAND[25]) is a software component that enables developers to integrate connected devices (e.g., IoT sensors and domotics) into their application.

A developer can use DYAMAND:

- embedded in his application.
- as a standalone system in the same network.
- via the remote management for distributed applications.
- as a testing framework to test their product or service.

A developer should use DYAMAND because

- it cuts down integration and testing costs.
- it works out-of-the-box.
- it enables them to focus on the business logic.

7.1.3 IDLab IoT data layer

IDLab's IoT data layer[26] is a highly scalable microservices based platform capable of reaching industry grade throughputs (simple setups can cope with way over 100000 messages per second). It is

[25] https://confluence.dyamand.ilabt.imec.be/display/HOM/Welcome
[26] https://idlab-iot.tengu.io/api/v1/docs/

especially tailored to integrate any sensor device and store its data as timeseries. These timeseries are served back to users or client apps in a very efficient paginated RESTfull API. The API itself is secured via OAuth and OpenID Connect standards, shielding and scoping the data to guide over the privacy aspects of the data in the data layer. The IDLab IoT data layer is managed by Tengu and served on IDLab's VMWare data center, although it can be fully instantiated on many different cloud providers.

7.1.4 JFed

jFed[27] makes it possible to learn the testbed federation architecture, workflows and APIs, and makes it also easy to develop java based client tools for testbed federation. The suite is built around the low level library, which implements the client side for all the supported APIs; and a high level library, which manages and keeps track of the lifecycle of an experiment. On top of these libraries various components were developed to allow thorough examination and testing of these APIs, as well as an user-friendly graphical Experimenter GUI to allow end-users to use the testbeds.
The most important components are:

- *jFed Experimenter GUI* allows end-users to provision and manage experiments.
- *jFed Probe* assists testbed developers in testing their API implementations

jFed Automated tester performs extensive full-automated tests of the testbed APIs, in which the complete workflow of an experiment is followed. This tool is used as part of the Fed4FIRE testbed monitor.

7.1.5 Tree Models

Tree models are a popular class of methods in data mining that employ interpretable rules to solve classification and regression tasks. Tree models provide separation between the two classes (class 1: failure, class 2: non-failure) and then iteratively grow the tree by selecting further features until certain stopping criteria are met. To select the optimal features, it is common to employ information theory measures, such as the information gain. Moreover, certain pruning techniques have been developed so as not to allow the tree model to overfit the input training data. Recent development in tree models that prevents the model from overfitting are Random Forests, which train multiple decision trees and make prediction based on averaging among different learned tree models.

7.1.6 Unsupervised Learning

Unsupervised learning is a class of methods that aim to identify the hidden structure of unlabelled data. In case of lack of data on whether a failure has happened or not, a clustering method is used. In this method the goal is to identify groups in our data such that the elements in each group are maximally similar to each other and dissimilar to the elements of other groups. There are several

[27] https://jfed.ilabt.imec.be/

algorithms for performing clustering and they can be categorized based on the objective function that is employed to characterize the optimal clustering structure (k-means, normalized cut, ratio cut, etc.), the type of clustering structure output (hierarchical, flat) and the type of method used (i.e. based on probabilistic model such as LDA or based on objective function optimization).

Another unsupervised method that is used is Frequent pattern mining. In its basic form, frequent pattern mining algorithm receives a set of transactions (item tuples) as input and aim in identifying frequent patterns (subgroups of tuples) that can be found within these transactions. Frequent pattern mining is used as a main component in association rule learning and sequential pattern mining.

7.1.7 Statistical Learning

Statistical Learning is a framework of machine learning that deals with building predictive functions. The Linear Models and their derivatives are one of the simplest and commonly used statistical models. General Linear Model can model also nonlinear dependence between the input and the predicted data (e.g. polynomial and logarithmic regression) by non-linear transforms of the input data. Assuming there will be Gaussian noise in the measurements, the relation between the measurements and parameters can be found via least mean squares method.

7.1.8 Probabilistic Methods

A Bayesian network is a probabilistic graphical model that represents a number of random variables and their conditional dependencies. Bayesian networks describe how e.g. machine parameters like temperature or speed depend causally on underlying failure occurrences, either directly or via a number of steps. Subsequently, Bayesian network inference can be used to infer backwards what the underlying cause must have been.

7.1.9 Data Integration (ETL)

Before data about machine and sensor device logs is fully ready to be Loaded to Big Data File System, the data is (1) Extracted from different sources plus it is converted into single format and (2) Transformed which includes parsing, clean-up, changing data types, changing the structure of data, removing unwanted columns, merging columns by concatenation, removing or modifying invalid entries in the data, splitting columns, using multiple columns to get numerical results, validating, etc. Software tools can help in doing ETL faster and more efficiently.

7.1.10 Deep Learning

Deep Learning is currently the highest-impact area of progress in the machine learning field that grew out of the quest for artificial intelligence. Deep learning approaches make use of huge training sets and gradient-based optimization algorithms to adjust millions of parameters and train the complex (multi-layered) neural networks which is supported by parallel computing on GPUs. Such approaches require (1) High performance computing cluster, (2) Deep learning servers and (3) A big data infrastructure.

Multiple Deep learning methods exist. Artificial neural networks (ANNs) came on the scene of machine learning research as viable modeling techniques to approximate universal functions that can transform input data into output within a vector space, allowing for robust representation of features that may otherwise be overlooked or inaccurately estimated by manual engineering or less complex techniques. However, the inability to effectively minimize an error function (and improve the accuracy of the universal function approximation) with a small set of hidden layers has been a major drawback for ANNs. One of the solutions to this drawback is unsupervised pre-training to accurately estimate the error function using back propagation, a technique credited to Geoffrey Hinton from the University of Toronto.

Unsupervised pre-training is a learning procedure that could exploit multiple layers of feature detectors to derive an extensive feature representation from unlabeled data. There are two models of pre-training, *denoising auto-encoder* and *restricted Boltzmann machine* (RBM). The objective in learning each layer of feature detectors is to be able to reconstruct or model the activities of feature detectors (or raw inputs) in the layer below. By ‘pre-training’ several layers of progressively more complex feature detectors using this reconstruction objective, the weights of a deep network could be initialized to sensible values. A final layer of output units could then be added to the top of the network and the whole deep system could be fine-tuned using standard back-propagation. Back propagation helps to estimate the derivatives of error functions to compute accurate weights for features derived in each layer of the network. These models for pre-training work remarkably well for recognizing handwritten digits or for detecting pedestrians, especially when the amount of labeled data is very limited, which is a common problem in many domains, such as the manufacturing.

Besides introducing unsupervised pre-training, Hinton also advocates another important technique called dropout, a simple but really effective way to prevent deep neural networks from overfitting. Hinton et al. introduced dropout training[28] as a way to control overfitting by randomly omitting subsets of features at each iteration of a training procedure. More specifically, it randomly omits a certain percentage of the feature detectors on each training case. Dropout is a kind of noise injection scheme, which controls overfitting by artificially corrupting the training data. This training method falls into the broader category of learning methods that artificially corrupt training data to stabilize predictions. For generalized linear models, dropout performs a form of *adaptive regularization*, which

[28] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1), 1929-1958.

has been shown to be equivalent to a variant of $L2$ regularizer[29]. A counterpart approach to dropout is *Maxout*, a natural companion to dropout designed to both facilitate optimization by dropout and improve the accuracy of dropout's fast approximate model averaging technique.

Several applications of deep learning use feed-forward neural network architectures which learn to map a fixed-size input (e.g. an image) to a fixed-size output (e.g. a probability of por quality). To go from one layer to the next, a set of units compute a weighted sum of their inputs from the previous layer and pass the result through a non-linear function. Another deep learning pioneer, Yoshua Bengio, introduced the rectified linear unit (ReLU)[30], a new type of non-linear function used in each neuron also referred to as activation function or squash function). In recent times, the ReLU has gained notable popularity in the deep learning community. Compared to the traditional, smoother non-linearities used by neural nets, the ReLU typically learns much faster in networks with many layers, allowing training of a deep supervised network irrespective of unsupervised pre-training. Besides their work on almost all aspects of deep learning, Bengio's team also developed Theano[31], a Python-based symbolic computational Deep Learning tool.

Also worthy of note are the contributions to deep learning research from Yann Lecun's team at NYU and the Facebook AI Research lab. One of the most widely used neural networks in image, audio and video processing is the convolutional neural networks (ConvNets), designed by Yann Lecun in the 1990s. ConvNets transforms input data (e.g. image pixels) to feature maps with distribution of weights and approximation of the feature location. Subsequent layers in the ConvNets generate higher order features by combining connected features with identical weights within a plane, thus learning the input data at various resolutions[32]. Lecun and his team later developed Torch[33], a deep learning platform based on LuaJIT[34], a just-in-time compiler for Lua – a lightweight programming platform comparable to Python.

Besides the aforementioned contributions in North America, researchers in Europe have equally achieved remarkable strides in deep learning. Jürgen Schmidhuber invented the long short-term memory (LSTM) architecture to leverage long-range contextual features. He also contributed to the development of recurrent neural networks (RNNs), a type of neural networks that exploits contextual information via memory storage[35]. RNNs can recognize complex moving images and automatically

[29] Wager, S., Wang, S., & Liang, P. S. (2013). Dropout training as adaptive regularization. In Advances in Neural Information Processing Systems (pp. 351-359).

[30] Glorot, X., Bordes, A., & Bengio, Y. (2011). Deep sparse rectifier neural networks. In International Conference on Artificial Intelligence and Statistics (pp. 315-323).

[31] http://deeplearning.net/software/theano/theano.pdf

[32] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.

[33] http://torch.ch

[34] http://luajit.org/luajit.html

[35] Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on (pp. 6645-6649).

generate detailed captions for online photos and videos, and improve online services that translate from one language to another (machine translation).

7.2 Trial present scenario

As mentioned in deliverable 2.1, the main focus area of the PCL Pilot is the injection moulding manufacturing process. There are two main areas for plastic part making in Drachten, on which Boost 4.0 focusses on the department of 'components' at first. This department focusses mainly on smaller plastic components for the shaver head. The department deploys about 20 injection moulding machines.

7.2.1 The general process

Injection moulding is a mass-production process for rapidly producing many similar parts. Under high pressure and high temperature, molten plastic is injected in a mould (the negative of the product). The mould is filled and rapidly cooled. The cooled plastic parts are then ejected.

An injection moulding machine can be generally described by the following overview:

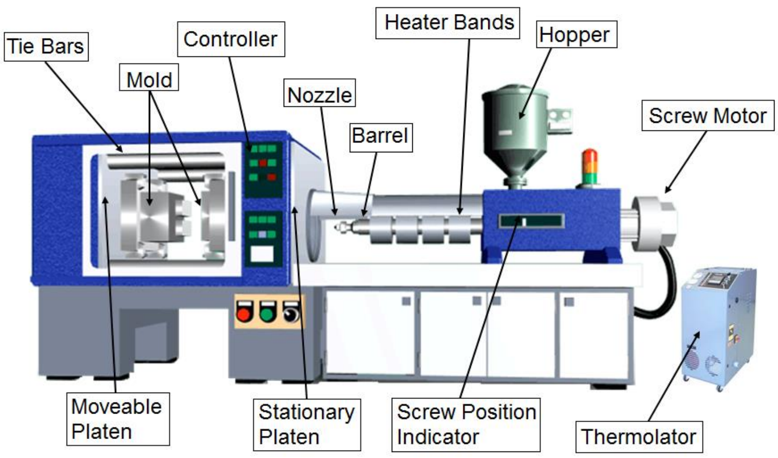


Figure 7-1 Most important parts of the injection moulding process (Courtesy of Autodesk MoldFlow)

Tie Bars – Support and align the platens, which then support the mold. The space between the tie-bars limits the size of the mold that can be placed in the injection molding machine.

Mold – Provides the formation and ejection of a molded part

Controller – Controls the injection molding machine

Nozzle – An adapter between injection unit and the mold which is designed to deliver the melt from the injection unit to the mold

Barrel – This contains the screw and is where the plastic is melted

Heater Bands – Heat the barrel and keep it at an appropriate, even temperature to prepare the melt.

Screw– This is internal to the barrel and is designed to meter the material from the feed to the nozzle and to help plasticate the polymer materials.

Screw Motor-Rotates the screw to plasticate and prepare the melt for injection into the mold.

Screw Position Indicator – Shows the size of the shot based on the position of the screw

Hopper– Plastic pellets are kept in the hopper where they are fed into the barrel

Moveable platen – Supports the B half (normally the ejection half) of the mold and acts to open and close the mold every molding cycle.

Stationary Platen – Supports the A half of the mold

Thermolator – Controls the supply and temperature of the coolant to the mold

7.2.2 The manufacturing process

The current manufacturing process is described by Figure 7-2. which is a schematic overview of the current process control loop. Notice that the current way-of-work focuses around manual intervention and control, and that data collection and advanced analytics are yet to be integrated.

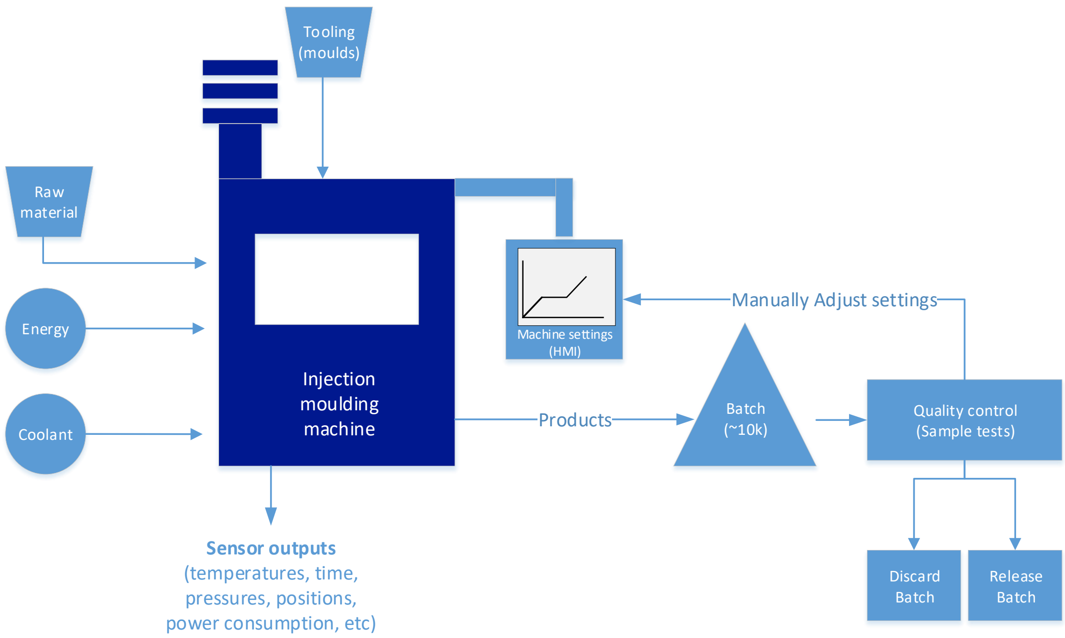


Figure 7-2 The manufacturing process

The main **inputs** for the injection moulding process are energy (to power the machine), coolant (to keep a stable temperature) and raw material (granulate). The raw material is delivered in batches, and fed into the hopper of the machine. The hopper is automatically filled from a container. If the container is getting empty, it will be switched out for a new material batch.

Tooling is another important aspect; the tools (moulds) are basically the negative shape of the product that needs to be made. Moulds can typically produce one to 16 parts at the same time. As the moulds wear (and are contaminated), the need to be maintained very well by a specialized maintenance department.

The **HMI** allows for basic monitoring and control of the machine. It usually consists of a touch-screen interface, and allows the operator to monitor sensor parameters and adjust settings. Usually, a machine also allows to set limits on certain parameters. The parameters are continuously monitored and automated action (usually rejection of the current part or stopping of the machine) can be triggered by the machine.

The output of the process are the end products. The products are temporarily stored in batches, from which a sample is taken. This sample is inspected and measured by the quality engineer or the operators. Based on the quality output, the machine can be adjusted (there are specific instruction for this) to improve quality. Note that one important aspect of **quality monitoring** for moulded parts are visual inspections; these inspections are based on norms and are difficult to 'digitize'. Quality is the most important metric for PCL.

Based on the final inspection of a batch, the whole batch can either be rejected (scrapped and recycled) or approved. After approval, the products are typically used in an assembly process.

The other aspect is **maintenance**, which can be broken down in two types:

- Maintenance of the machine: Reactive (react after break-down)
- Maintenance of the moulds: Preventive (react based on time interval)

Maintenance of the machine is taken care of by the maintenance engineering department (OTD). Maintenance of moulds takes place in a specialized workshop, sometimes with the help of an external supplier.

Production scheduling is order-based, depending on the production order put out by the internal customer (pull system via Kanban).

7.3 Weaknesses and bottlenecks

The main weakness of the current production system (Injection moulding) is the absence of data-driven decision making in practically all aspects of the production management system, despite the fact that data is available. This means that most decisions, like quality rejection, process control, maintenance actions, etc. are reactive or based on domain knowledge or experience.

This also implies that a data-driven culture is absent; one of the main challenges for adoption of Industry 4.0 related activities. One of the first steps will be to make data (and insights) available to the production stakeholders. Support from production management should also be provided, as without managerial support, it will be nearly impossible to develop solutions for production. This can be considered a bottleneck.

The data that is available should be used to benefit production performance. Basic items on a production level are monitored (production KPI's) like, cycle time, order progress, some process parameters, failure rates, etc., but these do not seem fully integrated in decision-making support. Action being taken need to be logged (preferably digitally) so analysis can be done on amount of actions, but also what the effect of these actions are. For example, why does an operator change a setting and what was the output of the action (did he react on quality issues, for example).

In order to successfully deploy solution across one (or more) department(s), a generalized interface should be provided. This interface needs to work with all available machines and interfaces. It needs to provide the operators and production managers comparable insights between injection moulding machines. The challenge here are the many different types of machines and interfaces; these need to be generalized in order to be effectively deployed. Based on the domain experience of the engineers in the Drachten facility, new decision support models need to be derived from the available data (note that despite the lack of data driven culture, there is much more data available than currently used). It is still unclear, however, if all this data is readily available or more data should be acquired.

This is always the case in new data projects, trying to discover what data is available and making sense out of the vast amounts of data. Somewhere in between these two steps, there needs to happen some data cleanup and the actual analytics to mold the data into actionable information. There is never a real guarantee the correct data points or data flows to actually create these higher-level insights can be combined. The heterogeneity of the different data sources is also something that can work against usable results. Here a good decision on the normalized data structure is going to be detrimental for a good outcome of the trials. Typically, two methodologies can be used: abstract data structure or canonical data structure. In the former, you check the common ground between what the different data sources can provide. The latter stores every single data element, even if there is no suitable counterpart for other data sources. Especially in a discovery phase the canonical data structure is the preferred choice, but this can put a real strain on the available resources. In plus, it makes the analysis and algorithms more difficult.

Next to this, the data is not free to be used, so it cannot be pushed to external clusters to do some deep number crunching. This makes it necessary to as efficiently as possible use the available server capacity, already present at the facility. Platforms and frameworks such as Tengu, can help in optimizing the available resources and turn any server into whatever you want it to be at any given moment (e.g. a highly distributable message broker, such as Kafka, or a Spark Cluster for big data analytics or simply a super performant Time Series Data Store).

A final point on potential weaknesses, as already mentioned before, feedback from real domain experts is always required to optimize the algorithms. It is also necessary to make the entire process, dynamic and in the end, self-learning. This means that in first phases of the trial the algorithms need to be guided by domain experts and in later stages the results of the algorithms need to be verified by the same domain experts. This a very time-consuming process, not only for developing these feedback mechanisms, but they also take up a lot of time from the workforce.

In order to develop reliable technologies with high accuracy that will be used to predict (1) quality of the molding machines' outputs and (2) unexpected failures of molding machines, the technologies have to be trained and tested on big data sets. The bottleneck could be that:

- the data set is not variable enough (e.g. number of occurrences with failures are limited). This leads to less reliable and accurate technology, as it cannot be tested for various scenarios that could happen after deployment of the technology.
- the data set is noisy. If the data set is distorted, the technology will be less accurate.
- The data is bad quality. This would lead to poor outputs of technologies.

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA	IMPACT IN THE COMPANY
Data driven culture	Due to the lack of decision-making based on data, the culture of Drachten might not be ready for data-driven decisions support	Management / Manufacturing	If decision cannot be made on data, they are made on experience of derived business rules. This might mean that the adoption of Boost 4.0 derived tools could be difficult. Both manufacturing and business should support data-driven production tools.
Logging of actions	Most actions are done manually, and are not logged digitally	Manufacturing	In order to build a data-driven supporting system (models, algorithms, etc.) we need to understand the action taken and the impact they have. The development and validation of results needs to be put against data, in this case, logs. On both levels, a digital log should be provided as soon as possible.
Production control system	All business rules that describe on how to run a manufacturing business (quality control, maintenance, orders, etc.)	Manufacturing	The current control system is hardly based on data-driven actions. This probably means that going to a data driven-system should allow for improvement of business performance.
Translation of domain knowledge	Domain knowledge about the processes and control systems need to be transformed	Management / Manufacturing	Transforming manual tasks to automated advanced decisions making models and algorithms need to be adopted. The models need to be based on the specific knowledge of system experts. The introduction of these systems will improve both the processes and the business.

Availably of (relevant) data in time	For all models, data need to be available	Manufacturing	The business needs to make sure that all relevant data is collected and stored, so it can be used to emulate and enhance existing business decisions. This is tightly related to the data-driven culture.
Impact on consumer value unclear	What is the impact on the end customers' product evaluation and what is the value	Marketing / Sales	For the business, it is not yet clear on how to relate the impact of the improved processes to the improved experience on customer level. While improvement can be tightly related to monetary value with respect to production costs, it should also be related to increasing customer value.
Platforms and systems	Since the Drachten location is vertically integrated, there are many different production systems integrated	Manufacturing	The large diversity of many different production systems, way-of-work, control aspects; but also technical differences like data formats and protocols make it difficult to get one integrated solution. Management of Drachten should support on standardizing the production control systems across the different departments.
Ground truth for improvement monitoring	In order to estimate production improvements, a ground trust should be established. This will be the baseline to measure improvements against.	Manufacturing	The Drachten site currently has a cost-model and a value stream available. It is, however, a one-time assessment and not driven by actual performance data. Establishing a baseline, build on current data, would provide Drachten to monitor production performance much better for all the different value streams.
Big data set	Not variable, noisy or bad quality data lead to poor outputs of predictive technologies.	Predictive technologies	Poor outputs of predictive technologies will lead to poor decision making after the deployment of technologies.

7.4 Trial future scenario

The first phase will solely rely on the infrastructure already being set up in Philips Drachten. Gradually and where it makes sense, specific technologies created or provided by the BOOST 4.0 consortium will be introduced. This to show interoperability, an important feature that the BOOST 4.0 projects tries to convey. Especially on the data transformation steps these specific BOOST 4.0 components can play an important role:

- From machine to the data backend: this will allow us to try out algorithms and easily integrate tools created by other BOOST 4.0 members, while using the PCL Drachten specific data.
- On the actual insights: in this respect PHILIPS might reuse visualization tools developed by other BOOST 4.0 contributors to present the insights/data to our domain experts.

Since the data (both raw data and derived information) cannot leave Philips Drachten, it is not possible to use BOOST 4.0 specific tools and methods to share information with external parties. This, however, does not mean that BOOST 4.0 mechanism cannot be instated to improve how data is integrated between different groups within the organization. For example, trying to make a generic platform to exchange the information within the plant can and should be based on BOOST 4.0 components, and thus should follow the principals of the BOOST 4.0 reference architecture (once concluded).

Predictive technologies will be part of 'IoT Hub' and they will communicate to 'Edge device' to show output on 'Real-time machine dashboard' and they will directly show output to factor operators on the 'Global Operations Dashboard'. The molding machines are the source of data. They will monitor different parameters of the machines and send the data set to 'Cloud' via secure connection. The data will be stored on 'Cloud'. The predictive technologies will be part of 'IoT Hub'. The technologies will collect the data from 'Cloud' and process the data for analysis. The output of the technologies will be shown to factory operators on the 'Real-time machine dashboard' and 'Global Operations dashboard'.

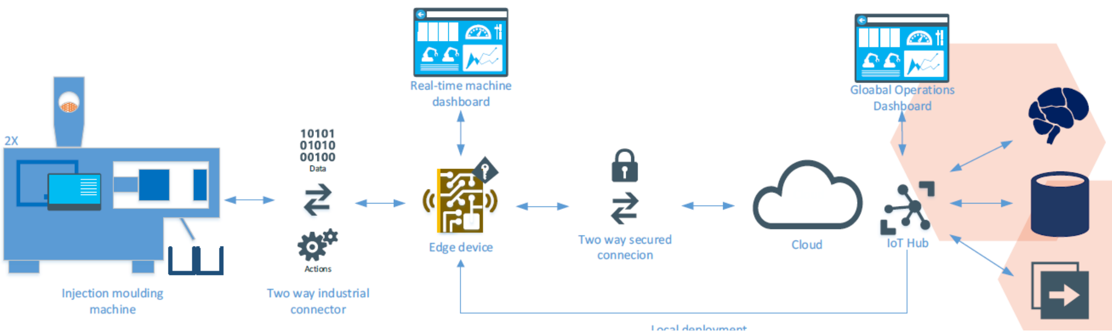


Figure 7-3 Future trial scenario

7.5 Expected results

One single platform for data collection, which support the many different communication protocols used by the injection moulding machines

A working method to 'translate' the original data structure of each machine into a normalized format. The ability to build generic models based the normalized formats and to deploy them on a data processing platform; the models can be re-used for the many different moulding machines. This generalization of manufacturing processes is not well described and seems to be challenging within manufacturing businesses.

The data processing platform should allow for streaming (real-time) analytics and be able to send out alarms and/or sent commands to the machine control to allow for closed-loop machine control. For this, specific requirements on response time and dependency on (network) connection should be taken into account. The generic models are based on current manufacturing decisions (domain experts) and/or based on newly gained insights by using data mining and advanced analytics, provided by our research partner(s).

The (real-time) data, as well as the output of the models shall be presented on a dashboard or HMI, developed specifically for injection moulding processes. The dashboards shall be generic and similar, independent of the type or brand of machine. Design of the dashboards shall be done together with the main users. This implies that multiple dashboards can be designed to provide stakeholders with the information they require to make data-driven decisions. These stakeholders include:

- Tool maintenance department
- Machine maintenance department
- Production management
- Production operators
- Quality management
- Manufacturing IT department

The models help to improve current production performance, related to the strategic goals and expected benefits:

1) Predictive Quality – 10% improvement on Fall Off Rate (FOR)
2) Predictive Maintenance – 5% less down time (OEE-A)
3) Intelligent process control – Contributes to predictive quality (in process-control)
4) Give operators better tools – Mean Time To Repair improvement of 5%

This indicated that effort should be spent on transforming the injection moulding organization to a data-driven organization. This means that the goals and vision is aligned with manufacturing, as well as taking the operators and shop-floor managers into account as main stakeholders. Any changes, after a certain stage in the project, should be well-communicated.

7.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 1: Plug and Play data connection and visualization					
BUSINESS PROCESS 1 - Injection Moulding data collection Platform					
1	Development of plug and play solution for setting up data in standardized way				
2	Set up two-way connection platform				
3	Validate new platform				
BUSINESS PROCESS 2 – Real-time Machine operator dashboard					
1	PoC Plug and play visualization				
BUSINESS SCENARIO 2: Process analysis insights & operationable actions					
BUSINESS PROCESS 1 – Big data IM process mining					
1	Development of predictive maintenance models				
2	Testing and scoring of predictive models				
3	Deployment of predictive models				
BUSINESS PROCESS 2 – Manufacturing support system					
1	Development of real time alerts				
2	Development of automatic machine adjustments				
3	Validate real time alerts				
4	Validate automatic machine adjustments				
GENERAL PROJECT MANAGEMENT					
1	Evaluation and KPIs collection				

8 Trial 7: GESTAMP automotive part prescriptive quality assurance factory 4.0

8.1 State of the art

8.1.1 Industry 4.0

Nowadays, the efficiency and sustainability of the manufacturing processes of high-tech products depend on the introduction of technologies towards digital, virtual and resource-efficient factories. In particular, the "European Factories of the Future Research Association" (EFFRA) identified in the "Multi-Annual Roadmap for Factories of the Future" for 2014-2020 the penetration of design and management manufacturing strategies and processes in the field of data collection, operation and planning, from real-time to long term optimisation approaches.

Industry 4.0, referred to as the "Fourth Industrial Revolution", also known as "smart manufacturing", "industrial internet" or "integrated industry", is currently a much-discussed topic that supposedly has the potential to affect entire industries by transforming the way goods are designed, manufactured, delivered and payed. It becomes apparent that the concept of Industry 4.0 still lacks a clear understanding and is not fully established in practice yet.

According to McKinsey article, most companies are trying to get better, the results tend to fall short: one-off initiatives in separate units that don't have a big enterprise-wide impact; adoption of the improvement method of the day, which almost invariably yields disappointing results; and programs that provide temporary gains but aren't sustainable.

They found that for companies to build value and provide compelling customer experiences at lower cost, they need to commit to a next-generation operating model. This operating model is a new way of running the organization that combines digital technologies and operations capabilities in an integrated, well-sequenced way to achieve step-change improvements in revenue, customer experience, and cost.

Many organizations have multiple independent initiatives underway to improve performance, usually housed within separate organizational groups (e.g. front and back office). This can make it easier to deliver incremental gains within individual units, but the overall impact is most often underwhelming and hard to sustain.[36]

[36] McKinsey on digital Services, Introducing the next-generation operating model

Currently, companies apply individual technologies, operations capabilities and approaches, obtaining only individual and local impact instead of a compound one. This is what is happening to Gestamp.

It is expected that the overall maintenance control and management system will be enriched with Product Quality information, i.e. by using online monitoring capabilities, so the quality of parts is correlated with the state of the machines.

The coordinate measuring machine (CMM) segment was expected to account for 67.8% of the global dimensional metrology market in the automotive industry by 2018. The figure below shows the expected trend from 2013-2018 for the metrology market in the automotive sector.[37]

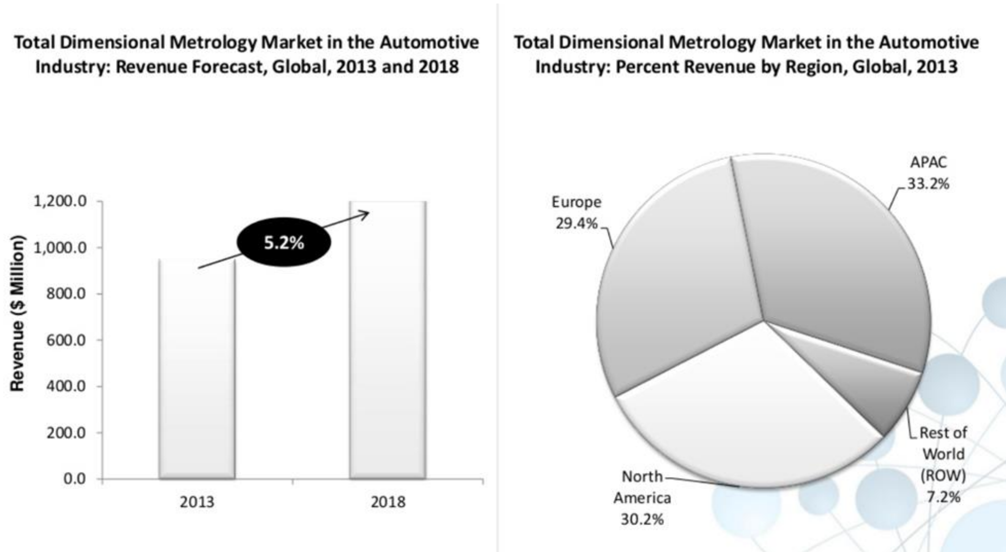


Figure 8-1 Dimensional Metrology Market in the Automotive Industry

For end users in developed countries, companies are looking forward to manufacturing high-accuracy and application-specific machines. As the need for enhanced safety, quality, and improved productivity grows in the automotive industry, particularly in the developed economies, metrology vendors' biggest challenge will not come from industry peers, but from meeting the insatiable needs of the automotive industry.

The market is likely to witness more consolidation, particularly in the inline metrology segment. Advanced fully-automated inline metrology systems will power the next-generation dimensional metrology solutions. Priority on improving existing inline metrology products and software packages will be the key demand. Customized software packages and cloud computing services are likely to become the most profitable business opportunities for dimensional metrology vendors.

[37] Analysis of the Dimensional Metrology Market in the Automotive Industry, 2013-2018. Frost & Sullivan, August 2014.

Moreover, smart factories will be measured on the level of intelligence and integration of production, assembling, quality inspection, and packaging. Quality inspection systems and reports generated using software will be integrated with servers using cloud computing. This study highlights TRIMEK M3 as the global reference solution for the automotive sector in the smart manufacturing domain for dimensional quality control in Industry 4.0.[37]

Cloud-computing adoption has been increasing rapidly, with cloud-specific spending expected to grow at more than six times the rate of general IT spending through 2020. While large organizations have successfully implemented specific software-as-a-service (SaaS) solutions or adopted a cloud-first strategy for new systems, many are struggling to get the full value of moving the bulk of their enterprise systems to the cloud.

The full value of cloud comes from approaching a holistic strategy to pursue digital transformation.[38]

TRIMEK metrological instrumentation provides optimum metrology solutions from the inspection and visualization of large parts, down to high accuracy 3D micro and nano-dimensional feature analysis; M3 Gages. These automated in-line scanning systems for reliable and efficient 3D information acquisition for multiple types of materials – metallic alloys, aluminium, invar, titanium, composites, thermoplastics …- can be directly integrated in the measurement process optimization platform M3®, M3 Software.

The M3 software is currently used for the organization, analysis and reporting operations of the metrology information. M3 Platform provides highly efficient, secure and flexible virtual part information management solutions for storage of massive 3D point cloud information and high-performance exchange and sharing of virtual part information.

INNO and TRIMEK have developed a global solution which modules will be extended and optimize during Gestamp trial. A global architecture from the digitalization of the part, crossing over an edge node to provide a physical web tool (M3 Workspace) and statistics or advanced analysis performed on the M3 Cloud. Thus, it is possible to take advantage of the storage and computational capabilities of the cloud to carry out advanced operations and provide smart added value services.

M3 Cloud is the name given to the platform aimed at integrating different data sources, mainly from inline inspection systems. The idea is that the M3 platform serves as a data repository as well as allows the interconnection of information from different sources using a format that can be used by other modules. The platform also provides computing mechanisms to process the data it contains and generate new information.

In summary, the development and extension of big data tools to detect and predict defects, providing valuable support for improved diagnosis is critical for the adoption of modern security posture,

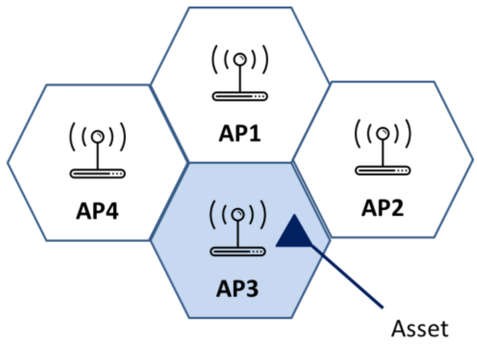
[38] N. Bommadevara, A. Del Miglio and S. Jansen, Cloud adoption to accelerate IT modernization, April 2018, McKinsey

working in an automated agile operating model, and leveraging new capabilities which will enable business growth.

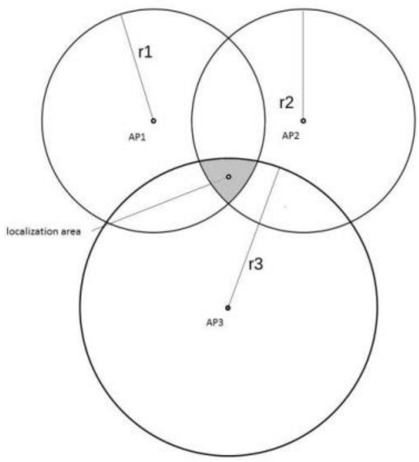
8.1.2 WI-FI-Based Positioning

Due to technological limitations of GPS alternative methods are needed to accurately locate assets inside of buildings. One of the most promising approaches is to use Wi-Fi to create what can be referred to as a Wi-Fi positioning system (WPS). The advantages of this system are that, in many cases, Wi-Fi access points already exist in many buildings. Furthermore, in some implementations of the technique, the user or asset does not need to be connected to the network for it to accurately locate them. An active Wi-Fi connection is sufficient.

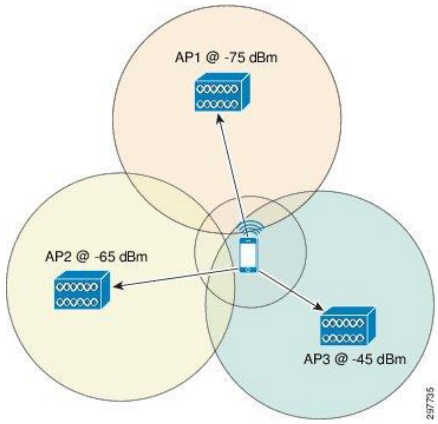
There are currently 3 main approaches to Wi-Fi Positioning:



- <u>Cell ID</u>: This approach is the most basic and least accurate. It consists of matching the target asset's unique ID to the known access points (AP) on the network. It is a relatively simply system to implement in that it does not require complex operations such as time synchronization or multiple APs. The main weakness of this approach is that is only accurate to the range of the AP. It is useful in situations where assets only need to be generally controlled and precise locations are not needed.
- <u>Trilateration</u>: This method involves the calculation of an asset position by determining the point of intersection of 3 circles. The position of each AP must be known for this technique to function properly. The trilateration method uses parameters of known WiFi networks like a frequency of Wi-Fi signal, its signal strength, the network MAC-addressee and real coordinates of Wi-Fi access points in the location. The received by mobile device signal strength can be used for distance estimation between the access point and mobile device. By using this method, one considers three or more access points allocated in the building. The signal strengths of these points are decreasing exponentially depends on distance between transmitter and receiver and random noise factor. Thus, this dependency can be considered as function of distance. The distance estimated by signal strength is presented as a circle with a radius around the access point. The intersection of three access point radiuses provides a point or an area of receiver.



- **RSSI**: This technique measures the distance from a Wi-Fi sensor to the AP by using signal-strength relationships. The goal of the approach is to map RSSI as a function of distance. This method requires a steep linear characterization curve in order to be properly implemented. Functions describing these curves are then used with live RSSI values as input to generate an (x,y) location prediction. Wi-Fi Fingerprinting then creates a radio map of a given area based on the RSSI data from several access points and generates a probability distribution of RSSI values for a given (x,y) location. Live RSSI values are then compared to the fingerprint to find the closest match and generate a predicted (x,y) location.



Precise location

- **Ultrawideband (UWB)** is a technology that was born during the 1960s, and whose name was coined by the Department of Defense of the United States in 1989. It was developed for radar, localization and communications applications. Its greatest attraction was that UWB's ability to operate in unfavourable signal-to-noise conditions prevented secure communications from being intercepted. Ultra-Wide Band, unlike other technologies, uses a radio pulse of below nanoseconds to transmit data in a wide range of bandwidth (normally above 500 MHz). Its transmission can be considered as background noise for other wireless technologies, therefore, in theory, it can use any spectrum without interfering with other users. It uses a small transmission power of -41.4dBm / MHz (which is limited by the FCC) which means that the power consumption is low. From here, the technology has been developed especially towards its use for the measurement of distances using the high precision method based on the time of flight (TOF).
- **VLC (Visual Light Communication)** the light fixtures emit positioning signals modulating the intensity of the light in time. Although communication through light has long been widely applied, only the high-power light used for general lighting has been possible modulate. Through the use of LED lighting systems (Light Emitting Diodes) obtain high energy efficiency, low cost and impact on the environment. The modulation of the LED light can be done at KHz frequencies, so that the VLC signals do not cause a flickering effect on the eye while transmitting information at speeds suitable for its positioning. This modulation is compatible with the majority of light switches currently available in the market. The signal emitted by each light fixture carries a unique identifier that distinguishes it from any other installation in the premises. The map of the locations of the lights and their identifiers must be created in advance and stored in the mobile system before starting to operate the system. An image sensor (camera) extracts VLC signals from an image that captures several sources of illumination (as many as there are in the field of vision of the sensor), demodulating and

decoding signals to uniquely identify. The process continues by analysing the relative and absolute distances of the emission points that determine the direction of arrival of each light. This allows the system to calculate its position with respect to each fix with an error of a few centimetres. Combining information about the relative position towards a lamp with information about its actual location, the system can find its position with an absolute precision of centimetres

- **Bluetooth (BLE - Bluetooth Low Energy)**: Bluetooth technology has been used extensively in connectivity applications for more than two decades. The application of geopositioning is based on measuring the attenuation of the electromagnetic signals emitted in the BLE protocol, with lower energy consumption. On the one hand, Bluetooth is very widespread at the user level through Smartphone applications iBeacon, Cornerstone etc. but in turn, in its positioning application, the erratic nature of the attenuation of the signal, makes it very vulnerable to the environment in which you want to deploy.
- **IoT platforms of geopositioning**: In addition to the technology for obtaining the positioning of tags, the importance of IoT platforms specialized in positioning must be highlighted. These software platforms ingest geo-positioning data either in internal hardware or in hardware outside the factory. The software facilitates data analytics in a first level as well as integration with other applications such as ERP, MES, etc.

8.1.3 Active Network Monitoring and Protection

The current market for network monitoring and protection is comprised of a series of threat and activity-specific solutions which are primarily based on signature or rule-based identification of threats. The available solutions have been developed with concrete objectives and to control a limited type of known threats. The market is now moving towards new approaches based on detecting unusual behaviour or traffic anomalies to protect against unknown threats. The following table presents a summary of the current state of network security solutions:

	Network security approach and solutions						
	SIGNATURE-BASED		RULE / BEHAVIOUR-BASED				ANOMALY DETECTION
Threat Type	Antivirus	Anti-Spyware	Firewall	IDS/IPS	SIEM	NTA (Network Traffic Analysis)	Threat detection based on ML
Known exploit	Best	Best	Best	Good	Good	Limited	Limited
0-day attack	Minimal/Null	Minimal/Null	Minimal/Null	Good	Good	Good	Best
Credential abuse	Minimal/Null	Minimal/Null	Minimal/Null	Limited	Limited	Good	Best

Mature | Maturing | In-devlopment

Minimal/Null | Limited | Good | Best

Figure 8-2 Comparative analysis of network security approaches and solutions

At the most general level there are 3 types of potential threats to network security: 1) Known exploits are those means of gaining access to a network or information which have been detected and analysed. These include known viruses and other types of malware. 2) 0-day attack/threat/exploit is previously unknown vulnerability is exploited to gain access. 3) Credential abuse is used to describe those cases where legitimate user credentials are used to gain access to the network.

To prevent these categories of attacks, there are currently 3 main approaches to network security which are often used in combination to provide as much protection as possible:

- Signature-based: These types of solutions work by comparing the unique signature of a file on the network with a database of known threats. These types of solutions are easy to deploy and offer automatic response capabilities. They are highly effective against known threats but are unable to stop threats which have not been classified. There is currently minimal development in this area as most efforts are concentrated on adding new threats to the known threats database.
- Rule/behaviour-based: Solutions which function by comparing the activity of a device, user or program on the network to set of predefined rules. These tools are able to respond to a broader set of potential threats as they do not rely on matching a unique signature but rather cover typical threat behaviours. For this reason, they are good at preventing 0-day attacks because they recognise a type of behaviour as potentially malicious. Their main weakness is they are not able to deal with behaviours which have not been classified, they cannot prevent attacks which use valid user credentials and they rely on manual coding of rules to be effective.
- Anomaly detection: This approach is fundamentally different from the signature or rule-based approaches in that it seeks to first establish what normal activity on a network is in order

to identify activities which deviate from this baseline. While theoretically not novel this approach is only feasible in real time with sufficient computing power and the aid of machine learning given the large amount of data which must be processed. The main difficulty with this approach is in the establishment of a baseline for normal activity and behaviour.

8.2 Trial present scenario

Gestamp is fully committed with Industry 4.0 as one of its main global objectives and is willing to develop new initiatives focused on the improvement of the overall efficiency of the plants. Gestamp wants to work on two of the main pillars, logistics and quality in order to test different solutions to optimize the logistics of the plant through the localization of assets and the inspection and quality control.

In manufacturing plants, logistic is a very important part of the process and very difficult to manage due to its complexity. The aim of the project is to improve current management of the logistics in a Gestamp plant by localizing in real-time the main moving assets such as forklifts, cranes and containers as some examples. This real-time information will allow to identify current inefficiencies and areas of improvement on the logistics and production processes. The expected improvements are, among others, as follows:

- OEE improvement by reducing downtimes in the production lines
- Anticipate production lines' needs
- Improvement on stock management
- Optimize movements of logistic means

It has been proven that industrial scenarios are very challenging when talking about indoor localization. For that reason, the trial of the project will be done in Gestamp Navarra plant, where many different production processes are present, and with a complex logistic that includes material flow between its two locations.

Gestamp Navarra plant is one of the plants that the group has in Spain), it's focused on Body-in-white (BIW) and chassis components production for the automotive sector, being its main customers VW, PSA, Volvo, Daimler and Renault.

Gestamp Navarra is located nearby Pamplona (5km), divided onto 2 sites: Orkoien and Salinas sites, it is specialized on welding process, cold stamping process, hydroforming process, laser cut process and shot peening process.

Welding process: Gestamp combines components of all our different manufacturing processes using welding technologies, using most advanced technologies such as metal inert gas (MIG) welding or metal active gas (MAG) welding, using parts resistant or spot welding in the welding of ultra-high strength steel and press hardening parts. Welding cells are typically automated and using automated robots to perform several of the most precise operations inside welding cells to achieve maximum cost reduction and ensure the highest quality assemblies productions.

Figure 8-3 Welding process

<u>Cold stamping process</u>: Cold forming technologies include forming operations in different types of machines. Sub-categories of cold forming include roll forming and hydro-forming. Cold forming allows us to manufacture a range of parts from small reinforcement parts to a complete car body side.

Cold forming involves the transformation of a sheet of metal at room temperature inside a forming die under pressure, operating various kinds of cold forming presses with different automation concepts with press forces ranging from 200 tons up to 2,500 tons. In order to achieve complex forms, parts must be pressed or stamped and cut in several steps, under different press technologies.

Figure 8-4 Cold stamping process

Depending on the size and shape of the part, the press process operations used to stamp the parts are chosen. Gestamp Navarra uses:

- Transfer presses for medium size parts with cupped shapes, material is moved inside the die by transfer bars in up to six operations. During the transfer press stamping process steel coil sheet is fed into a press and a blank is created where the material is cut from the coil strip. The blank is then pushed or transferred to the next station where the rough cup is created. Mechanical fingers transfer the cup to one or more subsequent draw stations until the rough, final shape has been created. The part is then transferred into additional stations that are used to establish critical diameters and lengths, features, and forms if required.

Figure 8-5 Transfer press

- Progressive presses for small size parts, where the material is always connected with the stamped part in the material strip and the finished part is separated from the strip after several forming and cutting operations. Progressive presses are mainly used for some deep draw stamping where the length to diameter ratio is low and part side features are not required. In progressive presses, the steel coil sheet is not cut, but is fed through the press. After several forming and cutting operations, and only once finished, the stamped part is separated from the material strip.

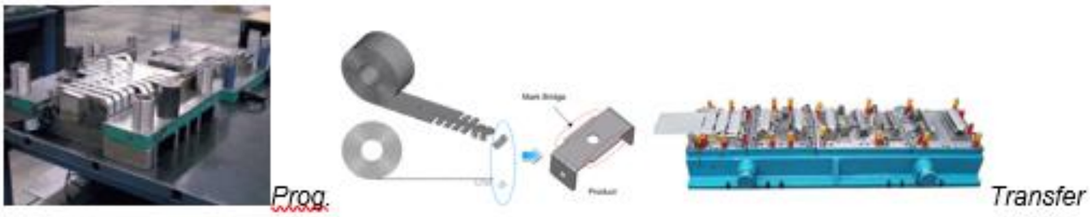


Figure 8-6 Process presses

Hydroforming process: It is a specialized type of cold forming that uses a high-pressure hydraulic fluid to press room temperature tubes into a die. The process consists of pre-bending a metallic tube and placing this pre-shaped tube inside a die with the desired cross sections and forms, and applying pressure to the inside of the tube held by the die. During the blowing or forming of the tube held in the die, holes can be pierced into the tube thereby avoiding secondary operations in most cases. Hydroforming allows complex shapes with concavities to be formed, which would be difficult or impossible with standard stamping.

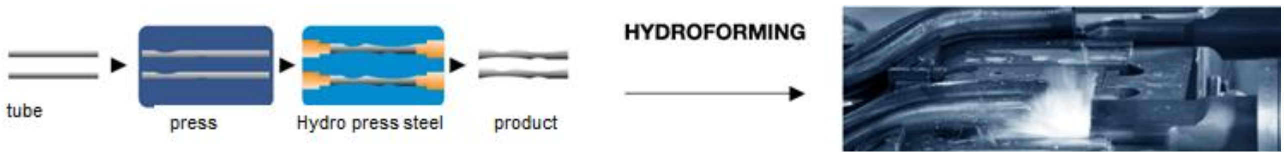


Figure 8-7 Hydroforming process

Hydroforming is considered to be a cost-effective way of shaping metal into lightweight, structurally stiff, complex and strong pieces. One of the advantages of using this process is that it enables us to create a three-dimensional tube that in cold stamping only could be manufactured by welding two shells together. The ability to deform thick materials makes this technology useful for chassis applications in particular.

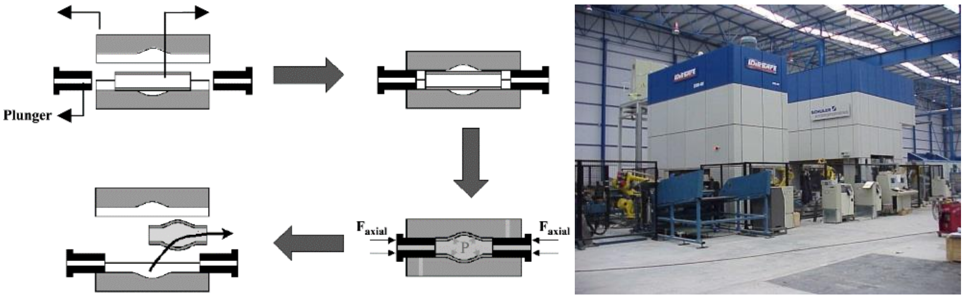


Figure 8-8 Process presses

<u>Shot peening process</u>: It is a cold working process used to produce a compressive residual stress layer and modify mechanical properties of metals parts. Shot peening is performed by accelerating spherical media toward the surface of a part (inside and outside areas). When the media hits the part, a small dent is formed, stretching the surface of the part. The material surrounding that dent resists this, and creates an area of compressive stress. When the surface of the part has these small dents all over the surface, there is a continuous layer of compressive stress on the surface of the part. This replaces the tensile stress on the surface with a compressive layer. The compressive layer stops the fatigue cracks and stress corrosion that typically start at the surface of the part.

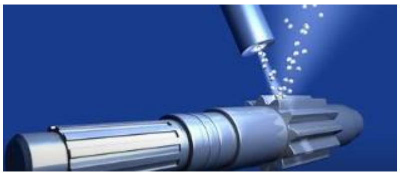


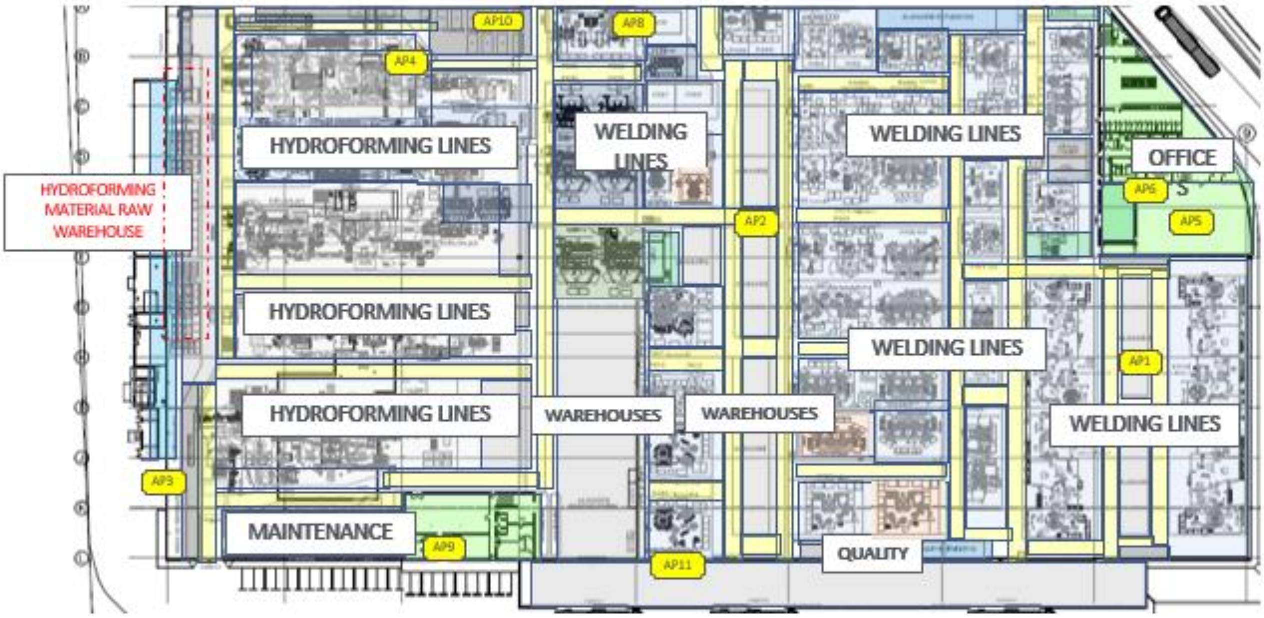
Figure 8-9 Shot peening

- ORKOIEN site

Located on the south of Pamplona, with an area of 53.000m² (27,000m² as buildings). This plant has conventional welding lines (Cells), hydroforming lines and laser cut for hydroforming products.

Figure 8-10 MN2 site

- General layout:
  Manufacturing lines, warehouses, offices and key areas (Maintenance and Quality facilities).



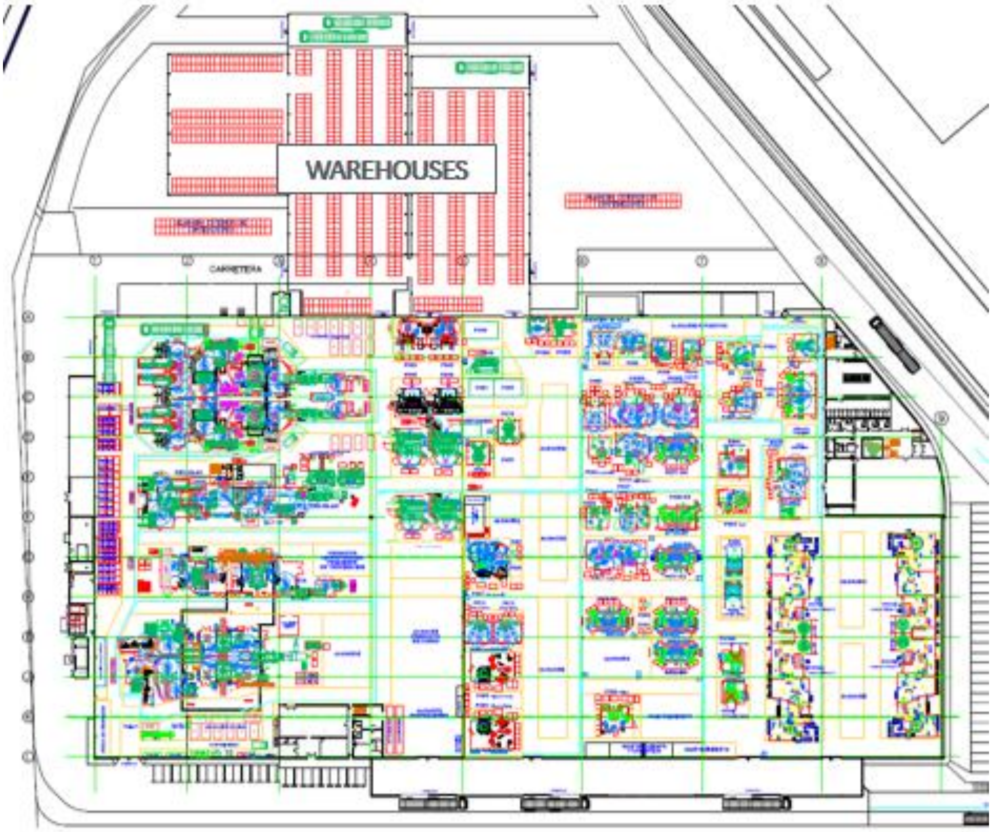


Figure 8-11 MN2 site layouts

- **SALINAS site**

West area of Pamplona, with 17,000m² of 56,000m² as buildings. Specialized on stamping, conventional welding and shot peening process.

- General layout:
  Manufacturing lines, warehouses, offices and key areas (Maintenance and Quality facilities).

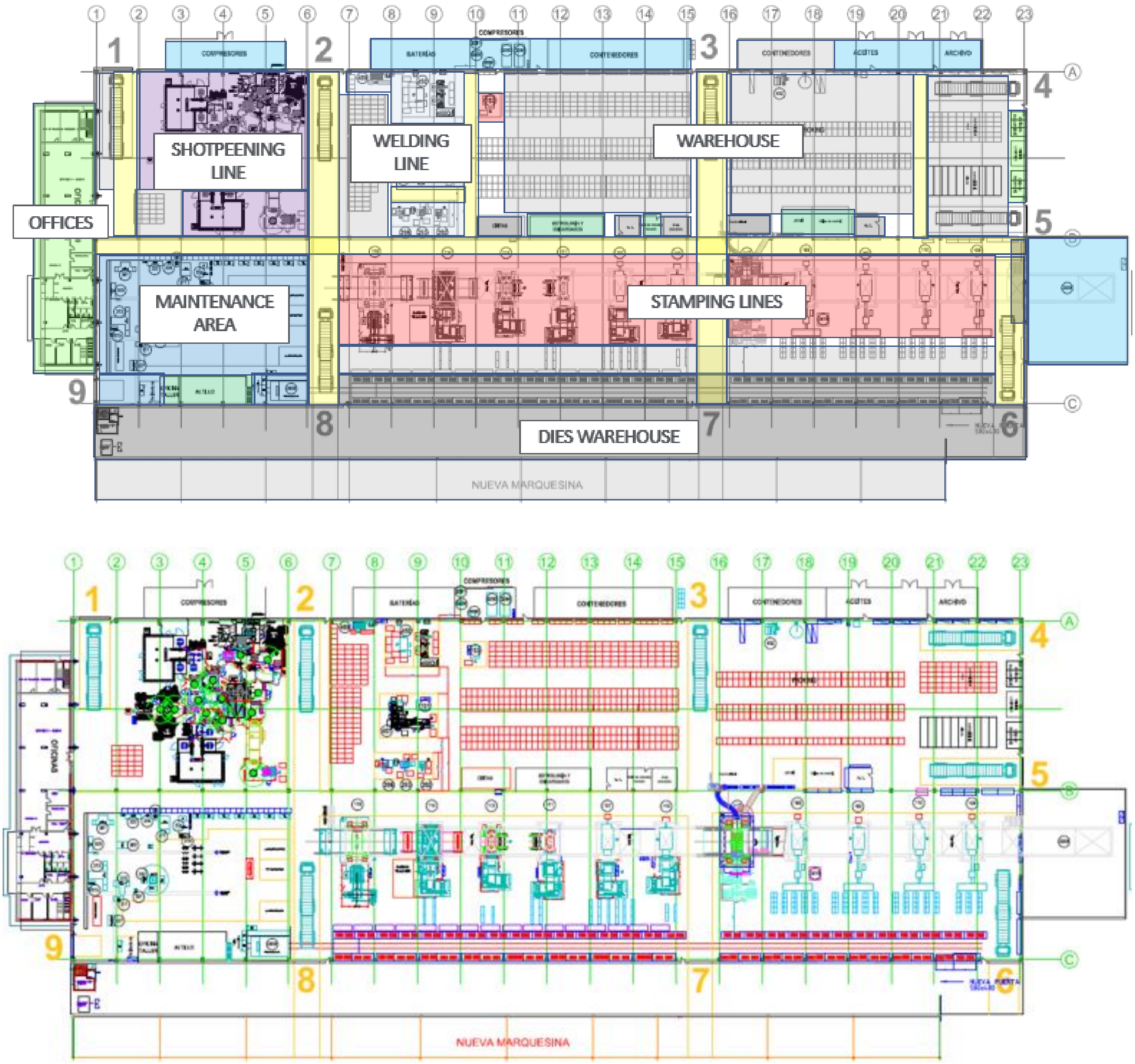


Figure 8-12 MN1 site layouts

The material flow is as follows:

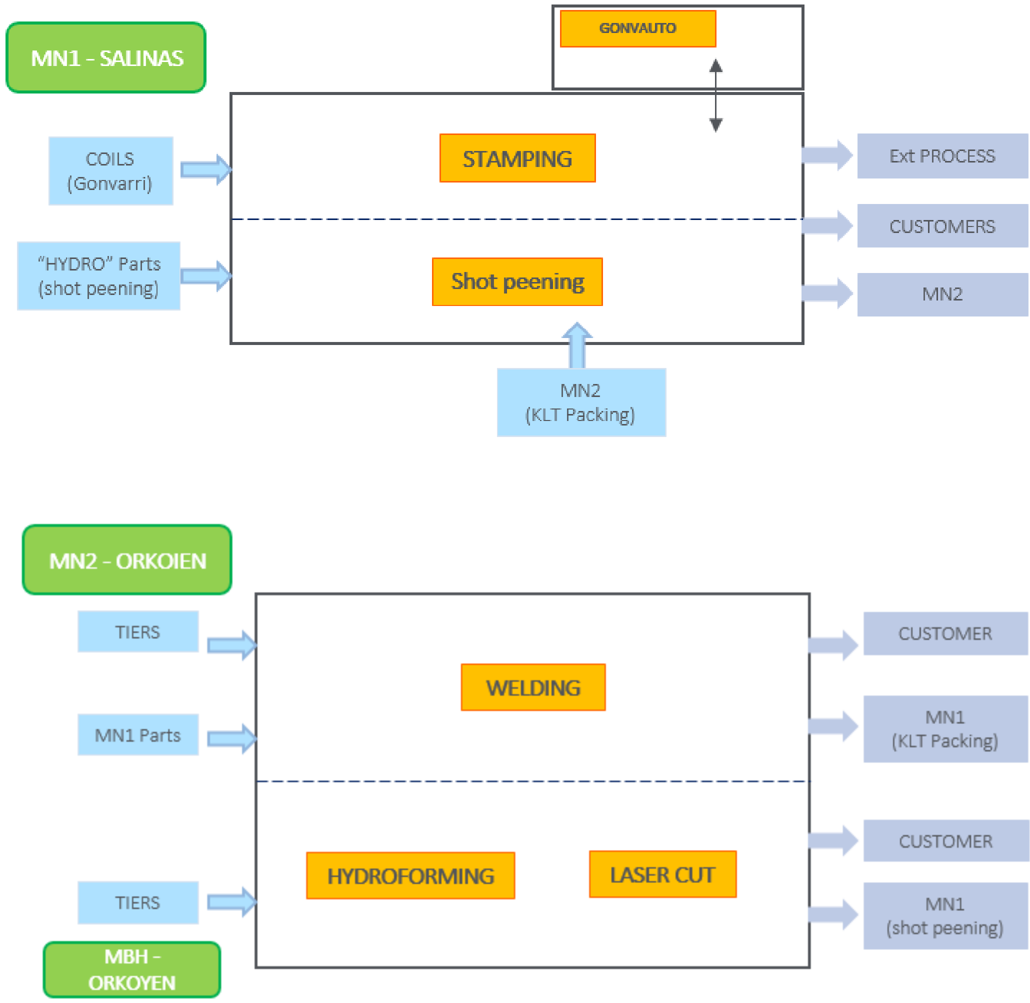


Figure 8-13 Material's flow chart

Regarding the inspection and quality control, Gestamp's plant process includes some quality check points after the key manufacturing lines (stamping and welding). The quality checks are carried out manually by blue collar workers who write the related quality information and decisions in notebooks/papers. Doubtlessly, the low accuracy of the checks especially after some of the technologies used in the plant such as welding whose conditions are more complex as well as the loss of information in relation to the quality of the manufactured products are major concerns in the factory. Thus, main problem focuses on the lack of quality information to take relevant decisions for the maintenance and system/process management.

One critical element is data access across factory information silos like CAPTOR MES or the ERP system and the interconnection and exchange of data among them to gain knowledge about the global status of the factory instead of the actual local and individualized view. Currently, in Gestamp's plant CAPTOR is gathering data about the manufacturing systems so it is possible to have information about the production process. On the other hand, SAP is being used independently for resource management. The big challenge will be the deployment of a trustfulness interoperable data network within the plant to allow automatic assistance and the visualization of plant's status to take decision at machine, process and factory level.

In addition, corrective and preventive actions, the approach followed in Gestamp's plant, have an individualized, partial and rigid approach which impacts in the low efficiency and sustainability from the economic and social point of view.

In spirit of this, two different experimentation environments have been selected to improve the efficiency of the manufacturing process: the Automotive Intelligence Center (AIC) and one of Gestamp's plant. On the one hand, the first site for the trial is the Automotive Intelligence Center (AIC) in Amorebieta-Etxano (Bizkaia - Spain); which is a unique value-generation centre for the automotive sector based on a concept of open manufacturing production innovation in which companies improve their competitiveness through cooperation.

At the AIC, the pilot will be developed at the Automotive Smart Factory (ASF). The ASF is a competence centre specialized in advanced manufacturing providing integral services for the implementation of industry 4.0.

Advanced Manufacturing supported by strong digitization and 4.0 technologies are key to transforming current competitive industrial companies into Top Performers. To reach this position, companies need to develop an individual strategic plan with an integral approach and efficiently evaluate the scope of the implications, with quick deployment of the strategy and rapid investment returns. The main focus of the ASF is to assist the industry throughout the entire process of transformation.

The ASF addresses advanced manufacturing challenges in an integrated way, developing ad-hoc manufacturing strategies for the industry, assessing the impact of implementing these strategies, providing specific or turn-key engineering services and advanced manufacturing training programs.

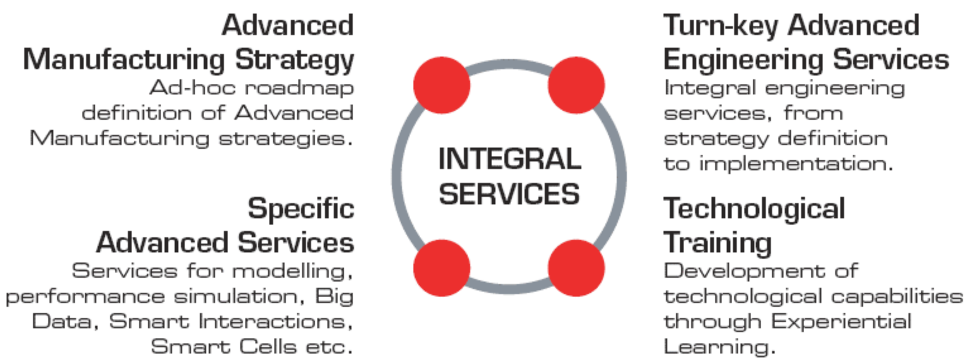


Figure 8-14 ASF services

The ASF has a combination of physical and virtual capabilities which makes it ideal for analyzing the benefits of these technological advances. The physical workspace includes cutting edge equipment, such as a stamping servo-press, an arc-welding cell and different control and in-line verification systems. It is also equipped with an AGV (Automated Guided Vehicle) to provide advanced flexibility in the handling process. The virtual workspace consists of a Smart System that manages the manufacturing process with an integral approach, and triggers real-time modifications of process parameters.

Figure 8-15 ASF

Specific technological topics that are physically tackled in the ASF include data mining, equipment and process monitoring, assets smart management, machine to machine communication, process simulation and control systems, digital quality management, human-machine interactions and new manufacturing training methods. All these forms the foundation of the factory of the future which seeks zero defect manufacturing and zero breakdowns, key to improving industrial competitiveness. To embrace these subjects the ASF relies on three main functions:

1. Manufacturing Intelligence
   - Advanced multi-variable modeling (big data).
   - Minimum variability in process.
   - Flexible in-line inspection systems.
   - Real-time deviation identification.
   - Simulation and optimization.
   - Advanced knowledge generation.
2. Track & Trace
   - Each part product and process data registration and analysis.
   - Physical and logical part identification.
   - Real-time modification of part-based process parameters.
3. Digital
   - Integrated plant level management system.
   - Automatic and real-time monitoring.

- Equipment advanced interconnectivity (IoT) for a proactive management.
- Visual factory for quick decision-making process.

Likewise, the Centre of Excellence in Zero Defects Manufacturing (ZDM) at the AIC, managed by Innovalia Metrology, will be part of the first part of the trial. The key pillars of this centre are: Industry 4.0, training, automation and digitalization. In this context, the main activities offered in the Centre of Excellence in ZDM are:

- Tailored technical support
- Research Infrastructures
- Technological tools and experimentation facilities
- Pilots and end-users access
- Access to technology and know-how
- Customised training
- Digital transformation consulting
- International networking
- Development of new products and new business models
- Dissemination and transfer of knowledge
- Technological development
- Market assessment
- Business advice
- Mentoring and coaching services

On the other hand, throughout the project, Gestamp will decide one specific plant to demonstrate the upgraded quality control process developed at the AIC installation. During the demonstration, quality and production data from different plants from Gestamp will be integrated and analysed. Thus, BOOST4.0 solution for Gestamp will not be a particular or specific solution for one plant but a global solution for any automotive manufacturing plant.

The present scenario of the pilot is depicted hereafter:

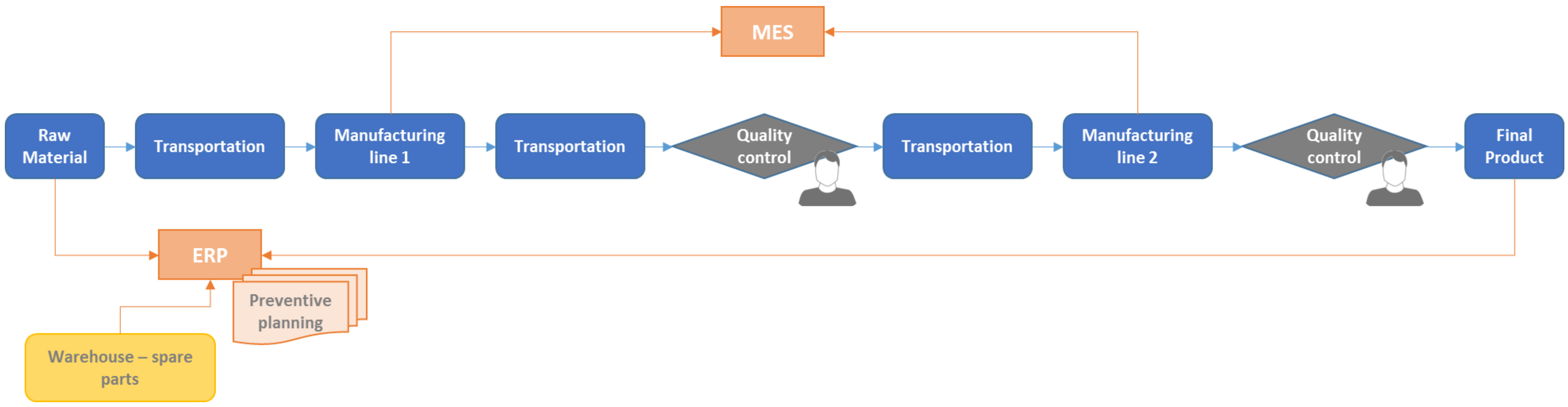


Figure 8-16. Present scenario diagram

The main improvements expected are as follows:

- Store quality information in a single database.
- Correlate quality information from different data sources in order to improve manufacturing process.
- Provide real-time information on manufacturing asset location and status.
- Establish framework for anomaly detection and real-time process monitoring.

8.3 Weaknesses and bottlenecks

The main constraint in the current production process is the lack of a global vision of the whole process which leads to inefficient processes in terms of energy, quality, logistics and maintenance. Information about machines, parts, products, assets status and behaviour are not always monitored and if so, the data is not intertwined. The decisions are made locally without having visibility of the real manufacturing situation (what is really happening in the factory). In short, there is an imperative urgency to implement frameworks for big data and data analytics in order to provide cognitive production planning and improve product and process orchestration.

Nowadays, Gestamp performs manual inspections using Go/NoGo approach according to the process quality verification strategy defined for the product characteristics. This is a manual process in which the outcome is just valid or not valid, and no information is stored for later purposes. In summary, this method leads to a loss of efficiency for the whole production process since not only the quality inspection is not accurate enough but the data of the quality control is not integrated and/or correlated to the process which is essential for the implementation of the most suitable actions to enable new business insights and achieve zero breakdowns and zero defects.

Thus, one of the main bottlenecks is the inability to connect and monitor large number of components, systems and devices and exchange information among them, so it is not possible to process and control the data generated. This bottleneck is also critical for the stock management. The connection and mixing of data from different sources and the implementation of full equipment and product availability strategy will permit to enrich the management of the stock, reducing the spare parts needed and planning the supply of the components in advanced so the cost of this item will decrease.

Moreover, the Navarra's plant needs to manage a huge amount of different assets for different activities. For example, forklifts are key elements in the factory to help to transport the raw material and product from/to the production lines. The ignorance about the movements and location of these elements are causing considerable loss of time for the process. For this reason, the use of customized positioning technologies could mitigate this weakness and increase the efficiency of overall factory.

As manufacturing profits rely heavily on maximizing the value of assets, asset performance gains can lead to big productivity improvements even in those cases where asset performance is only marginally improved. Furthermore, a reduction in asset breakdowns can reduce inefficiencies and prevent losses. Therefore, the primary goal of the detection of bottlenecks and weakness in the current scenario is to detect those areas of maintenance and continual asset performance optimization can be improved through data capture.

Machine logs contain data on asset performance. However, advances in sensor technology and the advent of the Internet of Things (IoT) adds a new dimension with connected assets and sensors capable of measuring, recording and transmitting performance in real time. While this data has enormous potential value, there is a clear weakness in the current ability of the company to process and analyse the sheer volume of incoming information. This results in the first major bottleneck of "Data Loss". This bottleneck is indicative of the combination of factors which leads to data not being captured, through lack of sensors, or not sufficiently analysed.

In summary, the big challenge of this trial will be the implementation of big data infrastructures and cognitive manufacturing strategies for a reliable interconnected ecosystem within the factory considering the heterogeneity of the equipment, component, and processes. The main needs identified are: condition monitoring at machine, process and plant level and secure data interconnection and analysis to support the decision-making process and improve the performance of the factory by reducing waste of material, energy consumption and defective products.

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA	IMPACT IN THE COMPANY
Local decisions	Data from the machines, devices, process and plant is not intertwined so the decision are making only with partial and limited information.	Management	To make decisions with limited information and in a local way only leads to poor efficient and highly cost processes.
Impossibility to store quality information, analyse it and correlate it to the production process.	Quality control is carried out manually by the blue-collar workers. Data from the quality check is not stored and, hence, it is not compared and correlated to other parameters of the manufacturing process.	Manufacturing	The quality inspection system currently used is not fully accurate and there is a loss of information in the supply chain so the number of defective parts increases. The importance of ensuring a good quality inspection and quality management from different production lines can allow Gestamp to achieve zero breakdown and zero defects. Consequently, there is a waste of money and reduction of productivity of the whole factory with the current strategy.
Lack of secure technology for big data analytics and interconnection of data from different data silos	Current technologies do not allow to acquire and integrate the generated data from the machines and sensors and real-time monitoring in a secure and reliable way. Moreover, there is not any cross-	Management and Technical support	The lack of big data analysis and collaboration between different business and plant systems causes poor performance and low efficiency at all levels (machine, process and plant). Less advanced and customized features and products.

	disciplinary collaboration between the plant systems and business applications.		
Corrective and preventive management.	These actions are focused on rectifying detected failures to restore the assets to the normal condition (corrective) or on general, regular and routine inspections to prevent breakdowns or malfunctions (preventive)	Management and Technical support	Considering these traditional strategies, actors cannot act in advanced and avoid malfunctions or breakdowns. Therefore, the prevention plans are not in real time without including the real behavior of the assets or the real needs of the production process. Subsequently, the economic losses and environmental impact are excessive.
Incapacity to locate and monitor assets such as forklifts and containers.	Nowadays, there are numerous assets which are needed to transport raw material and other elements within the indoor and outdoor area.	Manufacturing	Due to the lack of technologies to monitor and locate these assets, currently there are several mistakes and problems with their operation which cause loss of valuable time. In addition, there is not any related information about them such as their status or content.
Inability to plan stock needs in advanced and reduce spare parts in the warehouse	Spare parts are supply and acquired based on the suggestions/recommendations of the suppliers and without considering the real service or performance of the machines. In addition, the supply of these spare parts depends on suppliers' availability which	Warehousing	Currently, it is not possible to forecast the spare part consumption so it is necessary to have stock for every component, instead of acquire it when needed. This means high costs related to high amount of parts stored and, sometimes, obsolescence.

	sometimes might cause not planned delays.		

8.4 Trial future scenario

Gestamp will focus on improving the efficiency and performing of the whole plant located at Navarra in Spain. The improvement will consider the optimization of multiple aspects such as energy consumption, logistics, quality and maintenance. In this sense, the future trial will be based on the implementation of smart services to optimize these aspects within four different business processes which will be essential for the improvement of the overall efficiency of the factory.

The future scenario will implement Reactive, Coordination and Cognitive functions through smart services that will be deployed and operated over 3 levels: machine level, product level and plant level. The Reactive Functions are intended for fast reaction and close to machine operation (condition monitoring). The Cognitive Functions are intended for rich condition state and useful life remaining prediction as well as risk mitigation strategy identification exploiting reactive functions outputs, collective intelligence development through collaborative analytics and the plant level integrated information framework. The Coordination functions brings to live risk mitigation strategies through concrete dynamic maintenance plans and production control (operations) in the form of specific maintenance tasks to be executed on production machines and components.

As aforementioned, the main bottleneck is the lack of interconnected data among the heterogeneous systems used and the analysis of this data at machine, process and plant level. This bottleneck will be overcome by monitoring real-time data from the machines and collecting useful data from different data sources like MES, ERP…This data will be integrated, correlated and synchronized so all the departments and actors will be able to visualize and know what is happening during the process so the decision-making process will be facilitated and the optimization of the performance of the machines will be eased and ensured. One key activity to achieve this will be to setup secure and trusted data transactions and networks for connecting the OT & IT factory components.

During the trial, inline inspection solutions for quality control will be integrated in order to analyse and manage quality information about great number of components as well as finished and complex products (3D pointclouds of up to 10 million points) like car chassis frame. Moreover, this new approach will allow to integrate and manage quality data from different production lines to be subsequently analysed and processed, detecting defects in an early and accurate way, getting statistics, and customized reports so that all the quality information will be stored and managed in the M3 Cloud. In this context, this data will be accessible for the models and apps developed so will be able to be used in collaboration with data from other sources, preventing major failures and waste of material.

At the highest level, the overall objective of the trial will be to introduce new protocols in how factory data is used. Once the trial has been completed successfully the factory will have a fully integrated data collection and analysis system which provides information in real time to decision makers. The specific changes versus the current scenario will be concentrated in the following areas related to data gathering and analysis:

- **Data collection**: The main developments in this area will be the changes to how data is collected in the factory. This will include the implementation of new data collection methods including mobility sensors to track the location of assets in real time as well as the development of data collection protocols for existing sources of manufacturing information.

- **Data enrichment**: Once the data has been collected it will be enriched to facilitate its analysis as well as to increase the number of variables which can be measured in any given case. The enrichment process will involve automatically adding context information to the data collected from the sensors.

- **Data analysis:** The analysis of the captured data will be centred on solving the specific bottlenecks and weakness which have been previously identified as well as providing the company with a sufficient level of information on the manufacturing process to detect other problem areas. In contrast to the current scenario, this analysis will employ big data analytical techniques and be focused on improving process efficiency, reliability and velocity as well as proving predictive analysis based on historical trends.

- **Visualization / Actionable Intelligence**: The results of the data analysis will be the presentation of real time information on the current location and state of factory assets and processes in a manner which allows company decision makers to act accordingly. The major change from the current scenario will be the use of visual dashboards. The future scenario will allow to accelerate and improve the decision-making process encompassing actors from different departments and with different roles. Likewise, the trial will cope with the optimization of the quality control and logistics processes so the efficiency of the manufacturing process will be increased and zero-breakdowns and zero-defects achieved.

The following flow diagram that illustrates the whole process in the future:

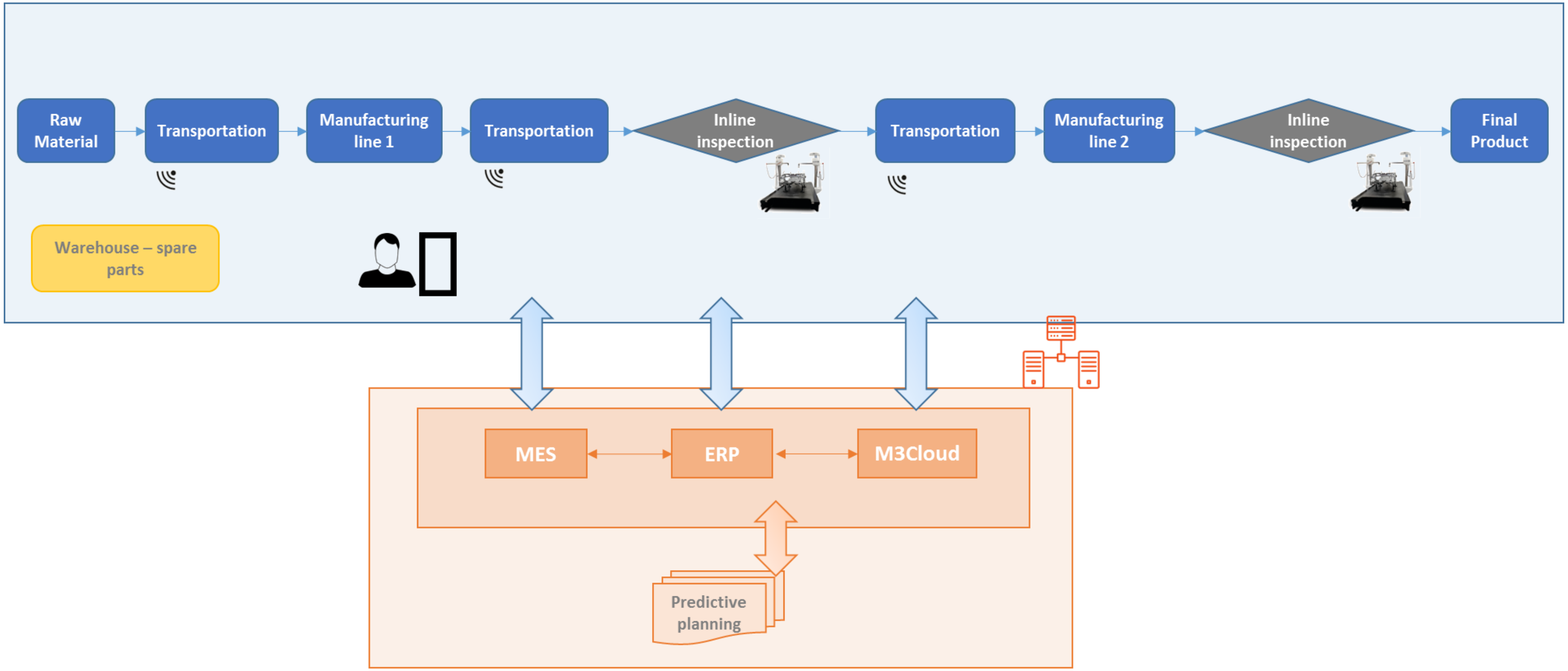


Figure 8-17 Future scenario diagram - General overview

The following diagram presents a general depiction of the data collection and analysis process following the trial. The main data sources will be the current data produced by the Manufacturing Execution System (MES) as well as the sensors, both new and existing, which will report directly from the factory assets:

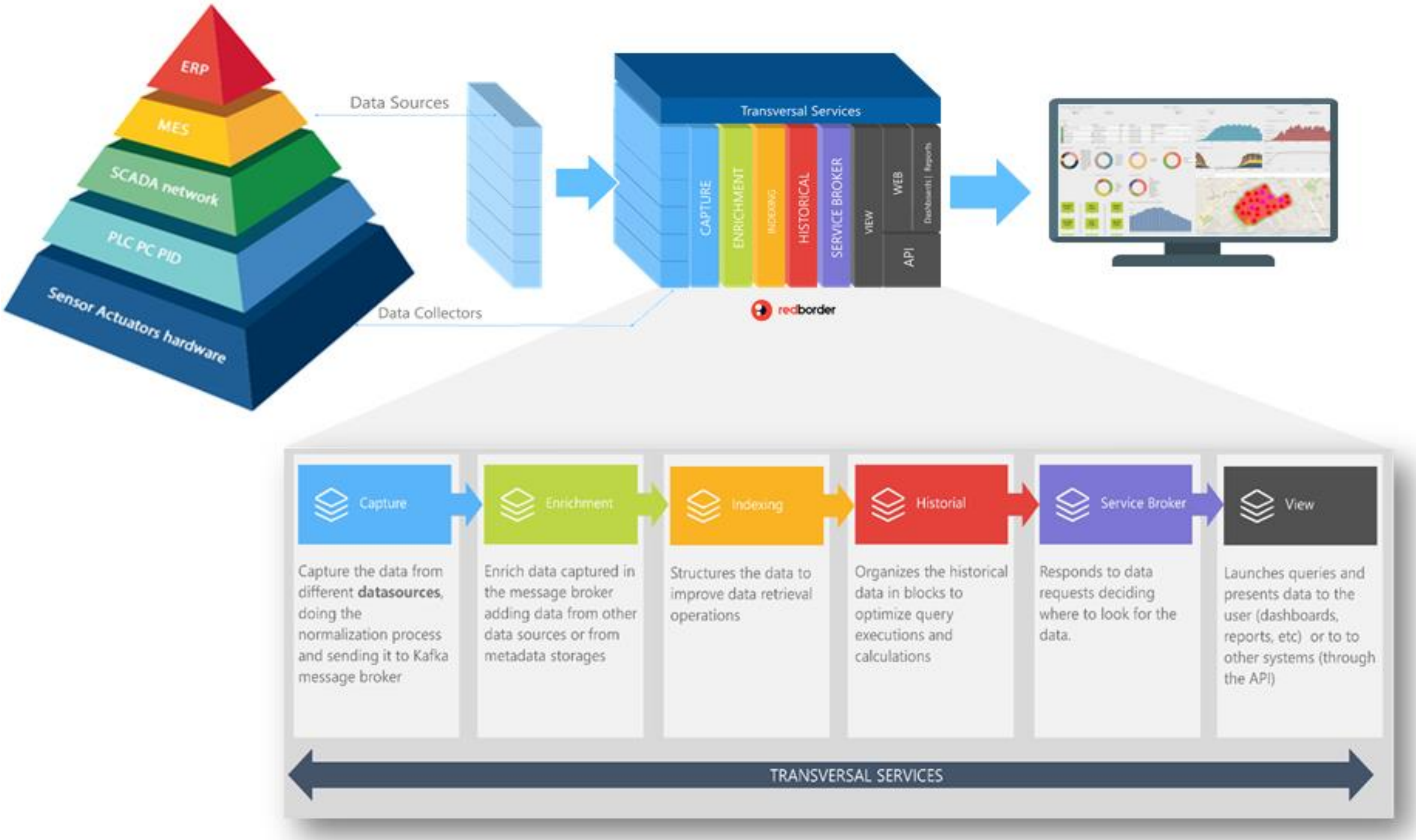


Figure 8-18: Future scenario diagram – Data Processing and Visualisation

The collected data will pass through the enrichment, indexing and historical organization processes which will be deployed, tested and validated in the trial. The service broker will coordinate data requests and present the results through the visualization module to the customized dashboard of each operator and decision maker at the factory. An example of the actual future deployment scenario is presented in the following diagram:

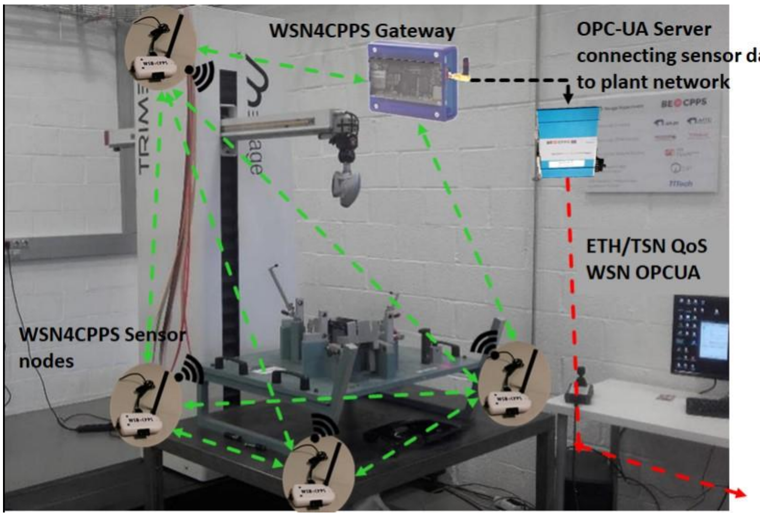


Figure 8-19: Future scenario diagram – Example Sensor Implementation

In this example the deployed sensors relay real time location information of the scanning device which is presented in a customized dashboard. The dashboard allows operators to determine the location of the assets in real time. The historical data analysis capabilities provide insights into how assets change location over time in the factory and provide insights into how efficiency can be improved.

Another smart solution implemented in the future scenario, is applied to inhouse logistics. More precisely to the handling of the materials around de factory: inbound raw materials, circulating semi-processed parts and finished products and tooling like matrixes and die cutters.

In both cases, they will rely on a mesh of UWB anchors located along all the facility. This mesh creates a space where any device with a UWB transmitter, can know their position in a precision of up to 10cm. Therefore, most mobile assets (forklifts, tow tractors, cranes, etc.) of the factory will be equipped with this such transmitters. Only by this implementation, all mobile assets will be monitored in real-time, pouring constantly data about their location. The collection of this data is essential to improve the efficiency of logistics operations, optimizing routes, uptime and size of the fleet.

On top of that UWB mesh and monitored mobile assets, the material monitoring solution is implemented. Materials are most of the time carried inside metal containers. The future scenario also monitors these containers in real time. This is achieved with an innovative solution that combines RFID tags stuck to the containers with RFID antennas in the mobile assets.

8.5 Expected results

8.5.1 GESTAMP Services

As described in *Deliverable D2.1 – Pilot Requirement and Use Cases Specification v1*, Gestamp is materializing different initiatives under Industry 4.0 and this "industrial evolution" with the aim of energy efficiency, zero defects, performance improvement, predictive maintenance and logistics.

In summary, with the combination of the existing digital sources and the new ones could be created (For instance, assets tracking), Gestamp expects to obtain:

- Production process optimization: With definition of an efficient data architecture, data correlation method and its understanding in real time. Current factory situation will be monitored, workshop performance and further surplus identified in order to enable the plant optimization.
- Reduce inefficiencies and further time losses due to no location assets: With the implementation of advanced technologies to locate assets in a real-time basis, the logistics processes will be optimized, gaining valuable knowledge about these assets.
- Analytical tools: Define key indicators and levers on the project to transform acquired data into added value to increase process efficiency, becoming more flexible and precise.
- In addition to the previous points, real-time assets location, data sources and visualization, it is expected to help operators to take decisions; show the right data on the right moment as well as to include particular alarms, events into the user interface for critical issues to focus on avoidance of product and process identification mistakes (At production and logistics areas), reduce uncertainly for the management and optimize the stocks.

8.5.2 TRIMEK

Development of 3D quality control data space and collaborative M3Cloud platform: Big data extensions of M3 software for collaborative analytics of quality information. Graph & multi-level visualization methods for individual 3D point clouds.

This outcome will overcome the bottleneck "Impossibility to store quality information, analyse it and correlate it to the production process".

8.5.3 ENEO

The expected results of the trial will be improvements in the future scenario with respect to the current scenario in each of the aforementioned data bottlenecks:

Data loss: The development of a rich and efficient data collection architecture and data correlation methods the future scenario will have a substantial increase in available factory data. The increase in available collected data will also be a result of the deployment of additional sensors to provide data on conditions which are not currently measured, including the location of factory assets. The result of the trial will be the definition of specific indicators to be analysed as well as new data collection protocols to ensure that valuable manufacturing, asset and network data is adequately collected.

Data delay: The changes to the data collection and analysis protocols will result in real time information on process and machine states becoming available to operators and decision makers. A concrete example of this will be the implementation of a set of alarms to alert operators when assets deviate from expected paths, processes take longer than expected and defects occur. These real time alerts will provide operators with the ability to make corrections to the process in a manner which greatly reduces the cost with respect the current scenario.

Process uncertainty: The overall outcome of the trial will be a significant reduction in the current information gaps which lead to overall process uncertainty. The improved data architecture will enable the creation of concrete, process-specific KPIs through which each aspect of the manufacturing process can be measured. These KPIs will specifically address areas such as process efficiency and rate of defects. The outcome will be the creation of specific action plans based on concrete conclusions from the data analysis to improve the overall process efficiency and ROI. Furthermore, the data analysis capabilities deployed and tested in the trial will enable additional improvements such as the implementation of predictive maintenance plans based on the historical analysis of machine performance.

8.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 2: Zero Defect Manufacturing powered by massive metrology					
BUSINESS PROCESS 1 - High-density metrology					
1	PoC Optimized massive quality data acquisition and high-performance processing and visualization tool (advanced colourmapping)				
2	Validation of the Optimized massive quality data acquisition and high-performance processing and visualization tool (advanced colourmapping)				

3	Trial of the high-density metrology process developed				
4	Evaluation and KPIs collection,				
BUSINESS PROCESS 2 - Virtual massive metrology					
1	Virtual massive metrology: programme execution (QIF execution), calculate Results (QIF Results), store of quality information on a collaborative environment, conduct advanced analysis.				
2	Development of data connectors and adapters				
3	Virtual massive metrology: generate QIF Model Based Design (CAD, PMI, geometries and tolerances, etc.), definition of the measuring plan (QIF Resources and QIF Plans), and definition of the digitalisation programme.				
4	Trial of the virtual massive metrology process developed				
5	Evaluation and KPIs collection,				

9 Trial 8: Volvo truck digital assembly factory 4.0

9.1 State of the art

9.1.1 Early detection – Predictive modeling

The anomaly detection problem, in its most general form, is not easy to solve and most of the existing anomaly detection techniques solve a specific formulation of the problem. A key aspect of any anomaly detection technique is the nature of the input data. The form and the nature of attributes of each data instances can determine the applicability of anomaly detection techniques. In general, most researchers have adopted techniques, such as machine learning, data mining, information theory, and statistics so as to deal with this crucial problem. In literature, one can find many approaches on anomaly detection such as classification [73] [74] [75] [76] [77] [78], clustering approach [73] [74] [75] [78], statistical approach [73] [74] [75] [78] [79] [80], and link analysis [81]. Anomaly detection techniques are applied in many areas such as wireless sensor networks [82], communication and social networks [76] [81] [82] and urban data [83].

New techniques, innovating algorithms and modifications of existing ones tailored to the needs and characteristics of the to-be-addressed use case, are utilized either on historical off-line or in real time data. These techniques will continually examine various conditions of the collected data, such as normality and linear trend maintenance through the calculation of linear trend profile of monitored features [84] [85]. The main objective of this process is to find possible deviations from normal conditions, detect for malfunctions and failures during industrial production processes and to point for alerts. For this part of detection, prior knowledge of the behavior of the data, will not take into account.

As for off-line and real-time predictive modeling, a series of known and popular machine learning techniques, such as Support Vector Machines [86], Decision Trees [87], Random Forest [88], Back-propagation network [89] and their boosting versions, such as Adaboost.*SAMME* algorithm for ensemble learning on two-class and multi-class classification scenarios [90] are commonly used in order to prevent future failures and abnormalities, to find hidden insights on various operations of the plant and to provide reliable decisions and results. A series of feature extraction techniques, such as Principal Component Analysis, Canonical Correspondence Analysis, Mutual information, is applied as a pre-processing step so as to find those features that contain the most useful information and will improve the performance of predictive models. Moreover, in order to understand (and trust) the predictive models, researchers use Local Interpretable Model-Agnostic Explanations (LIME), a technique that explains the predictions of any machine learning classifiers [91]. Artificial neural networks have been widely used in research and for practical applications since the early 80's. The evolutionary artificial neural networks utilize evolutionary algorithms like the most known Genetic Algorithm [92] so as to provide an alternative approach on process optimization. Deep learning techniques suitable for the form of existing data, such as fully-connected and convolutional neural networks [93] are the one-step-forward in machine learning, as an attempt to improve and enhance predictive performance.

9.1.2 Big data analytics

In modern manufacturing organizations, the challenge of collecting sufficient data for the efficient control of processes has shifted to analysing vast and continuously changing amounts of data for the extraction of useful information in real time [94]. A wide spectrum of sensors has been spread across industrial environments over the last decades causing the explosion of information in such environments in terms of volume, velocity, variety, and veracity; categorizing this issue under big data analytics. Unfortunately, most manufacturing companies do not make good use of all the generated and collected data [95]. Smart Manufacturing Systems (SMS) are foreseen as the solution to this through smart technologies implementing novel real-time control & data analytics solutions [96].

Past research in databases and information retrieval has focused on storage, search, and retrieval of information, functionalities that cannot cover the current need for the automated extraction of knowledge from big data resources [97]. On the basis of this gap, and in focus of the related issues and challenges, big data analytics has been continuously evolving and targets a variety of applications including manufacturing [98]. The analysis of data of sheer volume and dimensionality remains one of the main challenges in this field of research coupled with the fast data rates of modern measurement and assessment mechanisms.

Visual analytics has historically played a key role in business processes optimization. Existing tools [99] [100] [101] can be of great assistance for the visualization of spatiotemporal data in the shop floor, providing e.g.: temporal plots and heat maps indicating specific types of activities; representation of movement data joint with statistical analysis suitable to assist in discovering patterns and correlation; visualization of business processes detailing communication activities and summaries in addition to geospatial distribution and scheduling.

9.1.3 Blockchain

Manufacturing networks benefit from connectivity among customers, suppliers, manufacturers, partners, things, and cross geographic and regulatory boundaries. Wealth is generated as goods and services move across these networks. However, the growth of wealth can be constrained if the networks are heavily silo'd or inefficient. In today's manufacturing networks, each participant in a network keeps their own ledger[39](s) which are updated to represent business transactions as they occur. This is expensive due to duplication of effort and intermediaries adding a margin for services. It is clearly inefficient, as the business conditions for the transaction to occur – "the contract" – is duplicated by every network participant. It is also vulnerable because if a central system (e.g. Bank) is compromised

[39] A Ledger is the system of record for recording asset transfer in and out of a business.

due to an incident this affects the whole business network. Incidents can include fraud, cyber-attack or a simple mistake.

Blockchain is a digitally distributed ledger of a database of records, transactions, or executed events that are shared across the participants' parties in the business network. Each transaction in the system is time stamped and verified by a consensus algorithm.

Blockchain is a technology that supports a new generation of transactional applications and streamlined business processes by establishing the trust, accountability, and transparency that are essential to smart connected factory 4.0 operations.

Manufacturing is one of the sectors to be most impacted by Blockchain technology. According to a report from Frost and Sullivan[40], automotive ecosystem participants are expected to spend ~0.6% of their total IT spend on blockchain by 2025. Furthermore, smart manufacturing, supply chain logistics, retailing and leasing, connected living and IoT, mobility solutions, R&D, and aftermarket, are some of the automotive key functional areas to be based on blockchain technology.

Blockchain is a new technology that is driving significant transformation for business across multiple sectors, from financial services to supply chain and manufacturing to government initiatives and beyond. In the coming years, it is expected that the amount of applications will continue to grow as more understanding of the technology and its capabilities is gained. It is predicted that businesses and business models will emerge, based on smart contracts and blockchain efficiencies.

9.2 Trial present scenario

9.2.1 Workflow at Umeå plant

About 120 cabs are transported to Tuve from Umeå every day. Cabs leave the paint shop on a paint shop rack. They are transferred to a shipping rack at this point (this is owned by the paint shop) and a barcode that contains some form of ID information (Customer order number). The process is called undocking. The racks that the cabs are stored on, have a small barcode on them at this stage. They go through a buffer area into an initial trim stage, at this point the barcode is read and detailed barcodes showing chassis number are printed and applied. During the rest of the process, it is possible, but very unlikely, that a cab would be separated from the rack or that the rack would be replaced (e.g. rack becomes bent or broken.)

Cabs are stored in the yard in Umeå just on their racks. Cabs on racks are put on trailers for transport to the train station (Volvo dedicated area). Five trailers (dual articulated) with about nine cabs do eight journeys each day. At the station, cabs are removed and placed on the ground (for a short time). They are loaded into a wagon. A train is not supposed to be broken up, but it may be in exceptional situations

[40] Frost and Sullivan report K13A-18 "Blockchain Technology Revolutionizing Automotive Industry", 31 march 2017

- e.g. a wagon has a problem. There is a dedicated train each day for Gothenburg. The train company can, in theory, say where the train is but does not seem to let Volvo know if there was a problem with a wagon and it's been removed (three times in two weeks recently).

Returning racks are shipped in some form of stack (10 - 15 racks stacked) They are held in an external storage yard until needed, then they are brought into the rack/cab loading areas. Racks are then unstacked (with the appropriate type chosen as needed) and placed on a mechanical system that is used to marry the rack and cab.

The train from Umeå leaves in the afternoon to Gothenburg, the trip takes about 14 hours, it should arrive in Gothenburg the second day in the morning at 6:00 am. At the day when of a plant visit, it arrived at around 10:00 am. Cabs arrived at Gothenburg Arendal (there is a dedicated area for Volvo, the area is owned by the logistics company DFDS), they are first unloaded and then washed. After they have been washed, they are sequenced according to the information received from Tuve. The average time that the cabs stay at DFDS (truck company) is one to two days before they are transported to Tuve for final assembly.

9.2.2 Workflow at Tuve

The cabs are transported to Tuve with trailers. Five cabs are put on one trailer. They are unloaded into a Tuve yard area and then placed on a dolly for moving them around easily. No real need for precise location tracking here. Cabs on racks enter the trim line from the staging area through an inlet flow on their transport racks, where things like seats, microwaves etc. are added to the cab shell. There is their own production system that tracks each truck at different stages of the trim line so no need to track in that line, only on entry and exit. On finishing the trim line, some cabs ("Ikea" custom trucks, unclear) are removed from the metal transport placed on the wooden racks for transport elsewhere. The cabs on wooden racks and the empty racks exit through one door. Cabs still on transport racks (to be married to chassis at Tuve) exit through a different door. The in-door is on the right side about 10m from the "normal" exit door. "Normal" exit door is about 15m from the "Ikea" exit door. Exiting the trip area, the racks (with or without cab) are removed from their dollies

Finished cabs are moved towards the end of the plant on trailers (but removed from the trip area dollies) then placed in the final assembly storage yard (maybe 10 - 15 trailers) which is used as a buffer. This area seems to be an open area for the cabs (though there are some covered elements) which are placed on dollies to move them around in the final assembly stage. The final assembly dollies seem to be a different type from the ones in the trailer area. Cabs on their racks enter the final assembly area in an order that matched the chassis that are being built on the main line. There is a buffer of four or five cabs on racks, then the cabs are removed from the racks. The cabs have some final work done and are then taken to the assembly line and attached to the chassis. The now empty racks are returned to the final assembly storage area where that are stacked up and left (may or may not be under a cover).

The stacks (about ten racks high bound together) or racks are returned to the train yard three times a day, they are then loaded onto the wagons for a return to Umeå. Figure 3 and 4 illustrate flow diagram of the logistic process the scope of the Volvo use case.

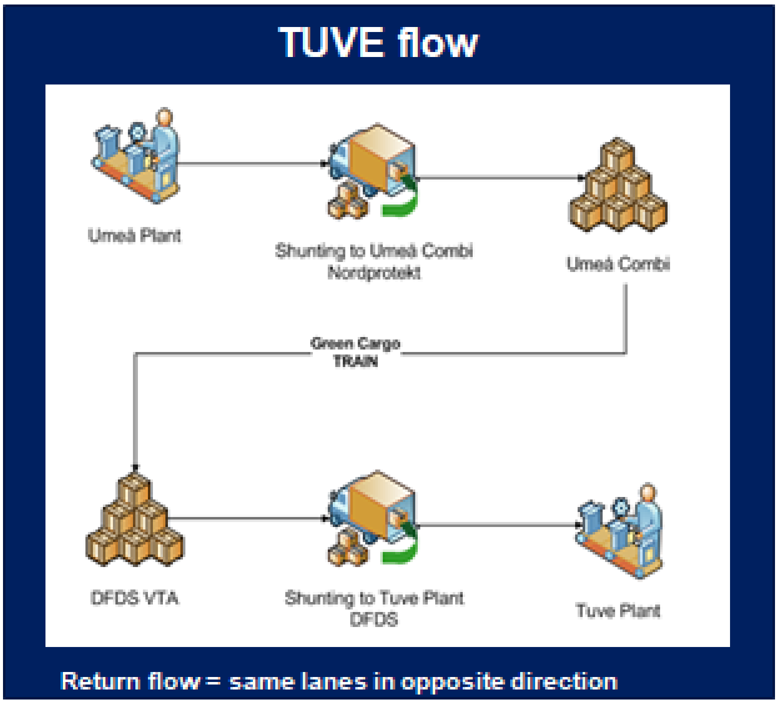


Figure 9-1 Flow diagram of the logistic process

9.3 Weaknesses and bottlenecks

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA	IMPACT IN THE COMPANY
Embracement of the technology by all parties in the network	Blockchain adoption depends that all parties in the network will use the technology	Management	If some of the parties don't join the Blockchain network the benefits of applying the technology are lost
Lack of real time data during cabs transportation	Without real time data the tracking during the transportation is unavailable	Management	During the transportation of cabs, delays and damage can occur. Cabs are transported on racks and every rack costs over 1500€
Data analytics and visualization tools are missing	Optimization of logistics planning needs data analytics and visualization tools	Management	Tools that can help the organization to optimize their planning of logistics and to provide decision support are missing

9.4 Trial future scenario

In general, a process change due to the Blockchain technology isn't expected. Blockchain will provide a more efficient process as it can help to track the cabs and racks and to provide real-time information about their status along with transparency and provenance to all parties in the chain (Volvo and external logistics company).

Data analytics and visual analytics tools will be considered as value-added services in this process. They will be offered as additional tools to the end users in order to enhance the decision support over the supply chain. The visualization of data and the data analysis will participate in the process in order to offer extra functionalities but the whole process will be able to be completed and without the use of these tools as well.

9.5 Expected results

9.5.1 Volvo and Chalmers

Expected results at processes level expected after the trial implementation

- A real-time tracking of cabs from the point where they leave the Umeå plant until they arrive to the Tuve plant
- The tracking of racks in the back-flow to the cab supplier
- Visualization of real-time data that provide decision support for the planning process
- Data analytics to learn from the data and identify patterns

9.5.2 IBM

Potential benefits of applying blockchain to the use case

Effectively management of the process by recording date, location, quantity, certification, and other relevant information. Blockchain enables the tracking of the cabs and racks from the manufacturing plant to the assembly plant. Blockchain can increase traceability, improve visibility, and compliance of the process. Blockchain provides all parties with access to the same information, potentially reducing communication or transfer data errors.

9.5.3 CERTH

Potential benefits of applying data analytics and visual analytics to the use case

- Early diagnosis of delays in cab transportation by applying data analytics methodologies and algorithms on both live tracking and historical data

- Enhancement of the decision support over the supply chain by applying algorithms such as a genetic algorithm. For example, an optimal proportion of cabs per routes will be available to the end users and will support their decision and planning.
- Visualization, analysis, and exploration of logistics data derived from multiple logistics sources will be offered to the end users from Interactive Visual Analytics tools. The multiple coordinated views of the data that allow a multifaceted perception and the discovery of hidden subtleties in it will enrich the decision support.

9.6 Execution plan

Step	Action	M1-6	M7-12	M13-18	M19-24	M25-30	M31-36
1	Define the goal						
2	Define the actors						
3	Define what is 'in' or 'out' of scope						
4	Define the elements of the use case						
5	Analyze available data						
6	Deployment of new sensors						
7	Extract Data from sensors						
8	Blockchain Development						
9	Development of analytics tool						
10	Development of cabs monitoring system						
11	Blockchain and analytics tools communication						
12	Application and testing of the forecasting mechanisms						
13	Development of possible value added services						
14	Overall system testing						
15	End user evaluation						

10 Trial 9: Whirlpool whitegoods spare part sensing customer service factory 4.0

10.1 State of the art

SAS provides different tools that can help business to achieve their needs. In the scope of the project, SAS will implement the functionalities described in WP8 with the following Tools:

1.1.1 SAS Viya

SAS Viya is a cloud-enabled, in-memory analytics engine that delivers everything you need for quick, accurate and consistent results. Elastic, scalable and fault-tolerant processing addresses complex analytical challenges and effortlessly scales to meet future needs.

SAS Viya provides:

- Faster processing for analytics.
- A standardized code base that supports programming in SAS and other languages, like Python, R, Java and Lua.
- Support for cloud, on-site or hybrid environments. It deploys seamlessly to any infrastructure or application ecosystem.

SAS Viya Capabilities

Cloud-enabled, elastic and scalable

SAS Viya is built to be elastic and scalable for both private and public clouds. Complex analytical, in-memory calculations are optimized for unconstrained environments and automatically adjust in constrained environments. The elastic processing effortlessly adapts to burst processing using available resources – scaling computing capacity as needed. This elasticity allows quickly experiment with different scenarios and apply more sophisticated approaches to increasing amounts and speeds of incoming data.

Open analytics coding environment

Data scientists and statisticians are provided with breadth of analytics capabilities that are easily available from the coding language of their choice. Whether it's SAS, Python, Java, R or Lua, analytical professionals can access the power of SAS for data manipulation, interactive data interrogations and advanced analytics. SAS Viya also includes public REST APIs to all underlying functionality, so software developers can add proven SAS Analytics to applications. And all analytical assets are united within a common environment to provide a single, managed inventory across your organization.

Fast, distributed in-memory processing

SAS Viya provides highly available, distributed processing crafted to handle multiple users and complex analytical workloads. Computing operations are automatically distributed across the cores of a single server or the nodes of a massive compute cluster. If a node fails, fault-tolerant processing ensures blazingly fast speed is retained.

All data, tables and objects are held in memory as long as required and for whomever is using them. Independent sessions ensure optimized processing for everyone. The SAS Viya engine provides a secure, scalable multiuser environment for concurrent access. Users can collaborate to simultaneously explore the same data, probe findings and identify analytically sound actions.

In addition, code written in the distributed environment is portable. It can be defined once, run anywhere and scaled to solve any size data problem. Code built on SAS Viya runs in stream, in database, in memory, in Hadoop, in public and private clouds, and even in device.

Resilient architecture with guaranteed failover

All analytical computations should run without interruption. The fault-tolerant design of SAS Viya automatically detects server failure, even across clusters. Processing is optimized and redistributed as needed. SAS Viya also manages several copies of data on the computing cluster. If a node in the cluster becomes unavailable or fails, the required data is retrieved from another block. These self-healing mechanisms ensure high availability for uninterrupted processing and automated recovery.

1.1.2 Visual Forecasting

SAS Visual Forecasting provides a resilient, distributed and optimized generic time series analysis scripting environment for cloud computing. This solution includes automatic forecast model generation, automatic variable and event selection, automatic parameter optimization, automatic model selection and automatic forecast generation. It also provides advanced support for time series analysis (time domain and frequency domain), time series decomposition, time series modelling, signal analysis and anomaly detection (for IoT). With SAS Visual Forecasting, you pick up the data once and run everything you need, taking advantage of in-memory, large-scale distributed processing. A scripting language optimizes and compiles your forecast based on where it is running.

SAS Visual Forecasting automatically analyses large numbers of time series so forecasters don't have to diagnose each series. The software determines the forecasting models that are most suitable for the historic data. When doing hierarchical forecasting, holdout samples can be specified so that forecasting models are selected not only by how well they fit past data, but how well they are likely to predict the future. An appropriate model is generated for each entity being forecast, based on user-defined criteria. Model parameters are automatically optimized. Any number of business drivers and events can be supplied and will be considered for inclusion in the models.

SAS Visual Forecasting provides a resilient, distributed and optimized generic time series analysis scripting environment. It supports fast, in-memory time series analysis. By nature, distributed systems break up large files and process each piece separately. This is problematic for time series analysis where the ordering of data is crucial. Time series analysis typically requires that time series data is stored contiguously in memory and in sorted order.

SAS Visual Forecasting shuffles the data so that each time series (or BY group) is copied into the memory of a single computing node. Each time series is executed on one thread of a node, and each node executes the compiled script for each of its assigned series. This makes time series analysis and forecasting possible on an enormous scale. And the scripting language is optimized and compiled for the machine it is running on, so users don't have to rewrite code for different machines.

Highly flexible forecast override

SAS Visual Forecasting adds a powerful, new capability that enables manual overrides to be made to a specific filter or group of time series defined by attributes, not just by hierarchical variables. For example, an analyst in the apparel Key Features industry may want to adjust a forecast for all products of a certain colour that is expected to be popular. Colour is not typically a level of the forecasting hierarchy. With the override capability, a custom filter can be defined of products meeting the colour attribute. Without this feature, if you wanted to apply an override to all products of a certain colour, you would have to manually enter the override to each product. Another example is sentiment, determined by text analysis of online reviews or user surveys. An analyst may want to increase (or decrease) forecasts for all products that have favourable (or unfavourable) sentiment. Creating filters saves a lot of time and manual effort when overriding non-hierarchical variables.

API support for working with open source:

While SAS Visual Forecasting has a broad range of forecasting models built-in, users can create their own customized models that perform well with their data. Also, with REST APIs, other applications can call SAS forecasting models.

Hierarchical reconciliation

Each series in the hierarchy is modelled and forecast individually. Forecasts are then reconciled at multiple levels of the hierarchy in a top-down fashion. Users can adjust a forecast at any level and apportion it to lower levels so the hierarchy maintains consistency, and individual forecasts (by products, locations, etc.) roll up to the top number. Without reconciliation, lower-level forecasts won't add up to the top-level forecast.

1.1.3 Visual Data Mining and Machine Learning

SAS Visual Data Mining and Machine Learning offers an exciting, new end-to-end visual environment that covers all aspects of machine learning and deep learning – from data access and data wrangling to sophisticated model building and deployment. In-memory, distributed processing handles large data and complex modelling, providing faster answers and efficient use of resources.

Flexible and approachable visual environment for analytics

Multiple users can currently analyse any amount of structured and unstructured data with the easy-to-use visual interface. Each project (goal) is defined by visual pipelines that break the analytics life cycle into a series of steps presented in a logical sequence. Branching can execute asynchronously. The visual interface (Model Studio) provides an integrated environment for the most common machine-learning steps: data prep, feature engineering, exploration, model building and deployment. Interactive tasks provide an easy way to apply sophisticated algorithms to large and complex data. These interactions also generate SAS code that can be save for later automation of tasks. In addition, code snippets and best practice templates are easily shared. Model Studio provides a highly collaborative environment for building, expanding and sharing models.

Highly scalable, in-memory analytical processing

This solution provides a secure, multiuser environment for concurrent access to data in memory. Data and analytical workloads operations are distributed across nodes, in parallel, and are multithreaded on each node for very fast speed. All data, tables and objects are held in memory as long as required, allowing for efficient processing. With built-in fault tolerance and memory management, advanced workflows can be applied to data, ensuring that processes always finish. You get dramatically reduced runtimes for large data and analytical processing, reduced network traffic and can take full advantage of modern, multicore architectures to find solutions much faster. The visual pipeline approach provides a collaborative, efficient environment for creating and deploying sophisticated machine-learning and deep learning models

Innovative statistical, data mining and machine-learning techniques

SAS Visual Data Mining and Machine Learning delivers an incredibly broad set of modern statistical, machine learning, deep learning and text analytics algorithms within a single environment. Analytical capabilities include clustering, different flavours of regression, random forests, gradient boosting models, support vector machines, natural language processing, topic detection and more. These powerful methods drive the identification of new patterns, trends and relationships between data attributes in structured and unstructured data. The solution also provides matrix factorization for building customized recommendation systems. With its ability to process high velocity and high-volume data sets, SAS Visual Data Mining and Machine Learning is uniquely suited for deep learning techniques. Deep learning algorithms include deep neural networks, convolution neural networks for image classification and recurrent neural networks for improved text analysis. Complex learning algorithms, such as neural networks, gradient boosting and random forest, can be automatically tuned for optimal performance, saving time and resources.

Integrated data preparation, exploration and feature engineering

To overcome usually time-consuming analytical data preparation activities, the drag-and-drop interface enables data engineers to quickly build and run transformations, augment data and join data within the integrated visual pipeline of activities. All actions are performed in memory to maintain a consistent data structure. Discover data issues and fix them with advanced analytical techniques. Quickly identify potential predictors, reduce the dimensions of large data sets and easily create new features from your original data.

Integrated text analytics

Designed with big data in mind, you can examine extremely large collections of text documents. Explore all of your textual data, not just a subset, to gain new insights about unknown themes and connections. Combining structured data with text data uncovers previously undetected relationships and adds even more predictive power to analytical models.

Model assessment and scoring

Test different modelling approaches in a single run and compare results of multiple supervised learning algorithms with standardized tests to quickly identify champion models. Then, operationalize analytics in distributed and traditional environments with automatically generated SAS score code.

Accessible and cloud-ready

Whether it's Python, R, Java or Lua, modelers and data scientists can access SAS capabilities from their preferred coding environment. And with SAS Viya REST APIs, you can add the power of SAS to other applications. You can also deploy SAS Visual Data Mining and Machine Learning where it makes the most sense for your organization: on-site, in a private cloud via technologies such as Cloud Foundry or in public clouds, including Amazon Web Services and Microsoft Azure. You can also access this software via the predeployed and preconfigured managed software-as-a-service offerings provided by SAS.

1.1.4 SAS Event Stream Process

SAS Event Stream Processing ingests large volumes of streaming data quickly – millions of events per second – so companies can understand events in the data while it's in motion. No data stream is too big or fast. It is possible to integrate, visualize, transform and analyze IoT data across the entire ecosystem – edge devices, data centers or the cloud. The solution's processing speed is bounded only by the hardware environment's limitations.

Incoming data is read through adapters and connectors, which are part of a publish-and-subscribe architecture. Event data publishes into a source window of an event stream processor. A visual interface makes it easy to define the windows, procedures and operators. In turn, it's simple to define continuous queries through which the data will stream. Streaming data is examined for patterns and can be intelligently filtered to store anomalies that demand deeper investigation. Or, if no relevancy is detected, the data can be discarded. Downstream applications subscribe to receive streaming analysis results with prescribed actions. This approach allows you to respond quickly to changing conditions and position your business for new IoT market opportunities.

For detailed information about SAS tools refer to:

- Visual Forecasting: https://www.sas.com/en_us/software/visual-forecasting.html
- Visual Data Mining and Machine Learning: https://www.sas.com/en_us/software/visual-data-mining-machine-learning.html
- Visual Analytics: https://www.sas.com/en_us/software/visual-analytics.html
- Event Stream Process: https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-event-stream-processing-106151.pdf

10.2Trial present scenario

Consumer Service Supply Chain Department is currently the main actor involved in the trial with the responsibility of managing the Inventory at European Level in order to maximise the availability of spare parts, accessories and cleaning products and reduce backorders and product exchange due to missing spare parts.

Here below a diagram to describe synthetically the main activities performed by the group (and impacted by the trial) and their responsibility all along the process:

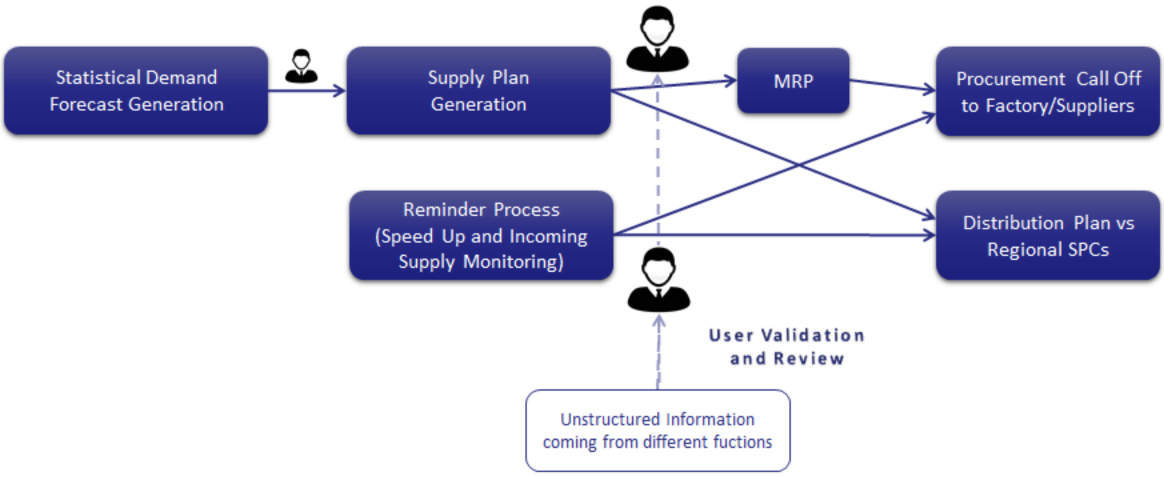


Figure 10-1 Present scenario representation

1) Statistical Demand Forecast Generation: this is a weekly process purely based on historical sales and consumption of spare parts in the whirlpool network. Granularity is currently per SKU / WAREHOUSE / MONTH as the historical demand is aggregated at that level. The Statistical forecast is currently generated by S099+ (Toolsgroup solution) and is then used for supply the supply planning process. The output review at that level is today near to zero due to the high number of spare parts to be evaluated the statistical forecast is usually accepted automatically.
2) Supply Plan Generation: this is a weekly process (or bi-weekly when needed) that calculate the required supply plan for each SKU / WAREHOUSE. This calculation is based on predefined target stock levels to reach, safety stock and demand forecast, suppliers constraints and total replenishment lead time, real on hand sales order when visible. Once the system creates the independent requirements, they can be reviewed though a monitor by the users in order to confirm the proposal or apply the necessary changes. In this phase a huge human effort is needed in order to evaluate code by code the system proposal and correct and integrate where needed with experience and information coming from other departments.
3) MRP (Material Requirements Planning): is then creating the final purchase requisitions at component level though BOM explosion in order to prepare for purchase order creation and supplier call off.
4) Procurement Call Off: this is the final creation of the PO the supplier/factory based on contracts and scheduling agreements.
5) Distribution Plan: this is the final creation of stock transfer orders for replenishments of the regional SPC.
6) Reminder Process: this is a weekly parallel process that is matching real time information like backorders, urgencies and suppliers feedback in order to send reminders to suppliers for

accelerating the incoming supply and/or react in next planning cycle to unexpected criticalities.

10.3 Weaknesses and bottlenecks

The main weakness in the current scenario is the delay in responding to unexpected events or trend changes in the spare part consumption. The visibility of what is really happening into the market is low and the information coming from other departments of the organization versus the planning team not structured. This is leading to an impossibility of managing in advance some important drivers for the generation of a good demand forecast and supply plan.

As the only element taken into account for automatic forecast generation today is the historical demand. A variation in a trend is captured by the model only after the things are happening that is in some cases too late to react. A good estimation of what is currently installed in every market is also missing. This is leading to the wrong assumptions that the historical demand patterns will remain the same in the future. A fix done on a defectiveness of an appliance by the factory should immediately be taken into consideration during the planning in order to rectify the trend of the consumption of the spares involved based on the estimation of the installed appliances with or without the defect.

The big number of active SKU to be managed every week by the planning team lead to the impossibility of controlling and enhancing manually the supply plan based on external variables or information transmitted to planning team in an unstructured way. The effort of the team to review and create every week a correct supply plan for the operations could be reduced.

The reliability of the suppliers is a key element in the calculation of the safety stock. We need to improve the reaction of the supply plan to adopt to the real service level of the suppliers that is currently calculated and monitored but not used during the supply plan generation. The impossibility to always rely on the declared performance lead to the need to increase safety stock in the warehouse.

The challenge is to create a forecasting model that can be used to integrate different type of data into the same database in order to give a good understanding of what is happening on the markets for planning purpose and react modifying the demand forecast accordingly (without waiting, like today, the things are happening).

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA	IMPACT IN THE COMPANY
Statistical Forecast generated only based on historical consumption of spares	Missing information during the spare part forecast generation lead to low quality forecast	Service Supply Chain	Impossibility to predict in a correct way the spare part consumption with high risk of obsolescence due to the high level of safety stock kept in the warehouse.
Relevant Information coming from other whirlpool function are not communicated to planning team or do not arrive on time	Missing information during the spare part forecast generation lead to low quality forecast	Service Supply Chain	Impossibility to predict in a correct way the spare part consumption with high risk of obsolescence due to the high level of safety stock kept in the warehouse.
The reliability of the spare parts component suppliers in not currently taken into consideration during the safety stock calculation	The safety stock is not taking into consideration the real service level of the suppliers, consequently WHIRLPOOL may risk backorder and /or overstock in a specific period	Service Supply Chain	Spares Parts missing availability lead to a consumer waiting for reparation and so dissatisfaction. The overstock is leading to high storage and obsolescence costs for the company.
Overload in the weekly review of the demand and supply plan	The planning team responsible for the stock of every active SKU in the network is currently reviewing all the system proposal to confirm the right supply generated every week	Service Supply Chain	The high number of SKU lead to the impossibility of the planning team to react immediately by changing the supply plan manually on every SKU impacted.

10.4Trial future scenario

During the trial the current scenario will be modified by substituting the current forecasting tool with a different technology for a subset of spares parts. The planning team will therefore have the possibility to run forecasting in 2 different tools (old one + new one) and thus the possibility also to compare it by measuring the forecast accuracy. For the selected parts the forecast generated by the new tool will be directly imported into the supply plan tool in order to materialise the predicted stock level needed in the warehouse with direct impact on service levels too.

The Boost 4.0 contribution is therefore mainly concentrated at the beginning of the process where we have the need to incorporate all the relevant big data produced across all phase of product lifecycle. The effort spent in the manual review of the forecast and the supply plan should decrease in favour of a more valuable activity of parametrisation and monitoring of the system performances. The remaining part of the process **will not change** compared as today: MRP/CALL OFF and REMAINDER processes will remain as is and they will use the same technology as today.

All the information collected and analysed by the system for forecasting purposes will be part of dedicated reporting and monitoring tool that can be used **by different departments** inside the organisation for different purposes. Not only the planning team but also **quality, manufacturing and markets** can have an immediate feeling of what is happening for a specific SKU now and in the future (in order to react before the thighs are happening). The initiation of a **predictive maintenance** can be also considered in this scenario for the smart appliances connected. In this case we'll be also able to match usage data and defectiveness found during the repair of the product in order to feedback product engineers.

The Reporting and Monitoring tool (directly connected with the elaboration of the big data for forecasting) will be also an important piece of the process and it'll be used by:

1) consumer service supply chain department in order to monitor system performances and analyse historical data versus the statistical forecast generated;

2) quality department in order to analyse defectiveness and give feedback to markets, factory and engineers on abnormal pick of claims.

Here below a graphic representation of the future scenario:

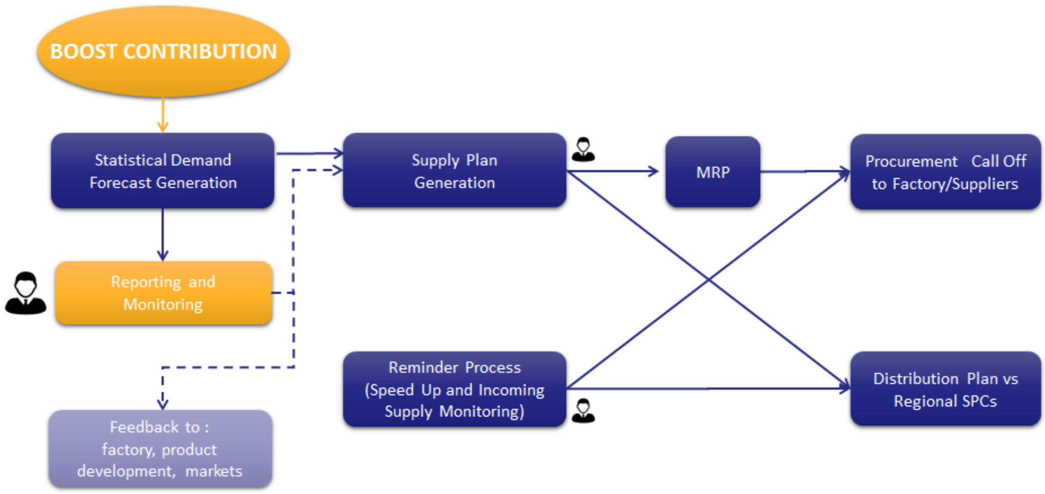


Figure 10-2 Future scenario representation

10.5 Expected results

	ACTIVITY	EXPECTED OUTCOME
Enlarged data Availability and Harmonization	The New Data Modelling will include all the relevant company data needed for planning in a more accurate way spare part	Big Data will be transformed into valuable information relevant for spare parts demand forecast generation
Immediate Event Recognition	Identify trends, relevant events, correlation of events or patterns, in order to generate set of information to be used by planning dept.	Things happening both at factory and consumer side will be immediately visible though an alert monitor to take the right decision
Demand Forecast Generation	Change the statistical engine in order to analyze all the information selected automatically in order to generate a more accurate sales forecast	Forecast Error Reduction
Workload Reduction of the planning team	Built a monitor that help the planning time to concentrate on the critical SKU management while improving the system capability to generate forecast save time of exception managements	Time saved to concentrate in other valuable activities
Model Implementation	Streamline the forecasting process	Automatically producing large-scale time series analysis and hierarchical forecasting with no human involvement
Model Implementation	Focus efforts on high-value situations	Forecast analysts don't have to build and monitor forecasting models for

		every time series. They can focus their efforts on strategic, high-value forecasts or problems that aren't suitable for automation
Model Implementation	Deliver forecasts that reflect reality	Automatically selection of the business drivers, holidays or events that aid in the forecasting process from variables supplied to the system. Ability to manually override forecasts based on groups that are defined using attributes instead of hierarchical variables
Supply Plan Generation	Trigger Supply Plan based on the new technique	Reduce the overall stock on spares that will became obsolete because not really needed and increase the service level by distributing in advance the necessary spare parts when/where needed

10.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 1: CONSUMER SERVICE SUPPLY CHAIN					
BUSINESS PROCESS 1 - Consumer Service Data Model Creation					
1	Data Source Analysis				
2	Data Extraction and ETL Development				
3	Central CS Data-Lake Loading				
BUSINESS PROCESS 2 - Spare Parts Demand Forecast Generation					
1	Data Model Creation and Software Customisation				
2	First Result Evaluation				
3	Spare Part Forecast Generation				
4	Evaluation and KPIs collection				
BUSINESS PROCESS 3 - Spare Parts Inventory Optimization					
1	Preliminary Results Evaluation				
2	New Inventory Planning Generation				

3	Evaluation and KPIs collection				
BUSINESS PROCESS 4 - Supply and Production Plan Optimization					
1	Preliminary Results Evaluation				
2	New Supply Plan Generation				
3	Evaluation and KPIs collection				
BUSINESS PROCESS 5 - Smart Connected Appliances					
1	Identify requirements and solutions				
2	Data Source Analysis and Extraction				
3	Data Model Enrichment and Report Preparation				
4	Evaluation and KPIs collection				
BUSINESS PROCESS 6 - Cross Department Quality Monitoring and Feedback					
1	Report Preparation and Data Understanding				
2	Feedback Process Implementation				
3	Evaluation and KPIs collection				

11 Trial 10: Benteler predictive factory 4.0

11.1 State of the art

The goal of predictive maintenance is to prevent unexpected equipment failures by continuously observing the status of the equipment and creating alerts in a proactive manner. Traditionally life usage (LU) models are widely used in reliability [102], mostly due to their ability to deal with small sample sizes and their flexibility to approximate a wide range of statistical distribution. In these models, the usage measures of the equipment are fit to a probability distribution (usually Weibull distribution [103]), to characterize failure behavior. However, recent research studies propose data-driven predictive techniques [104], that outperform the results of LU models. Due to the complexity of the problem a single data-driven technique cannot fully capture the whole process, hence a set of techniques need to be applied to enhance the robustness of the results.

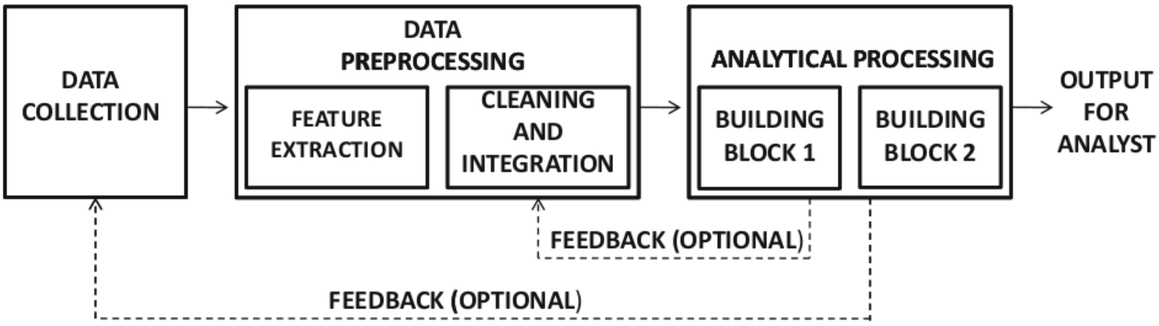


Figure 11-1 Predictive maintenance pipeline [105]

As depicted in the Figure borrowed from [105], predictive maintenance follows the same pipeline procedure as most of the data mining scenarios. There are three distinct phases (i.e. Data Collection, Data Preprocessing and Analytical Processing), where each exhibit its own special characteristics that have great impact on the design of the corresponding solution.

Considering the Data Collection process, data can have the form of continuous values (e.g. raw input from sensors in time series form), discrete event sequences (e.g. event logs from maintenance engineers), or a mixture of both, where continuous values are correlated to "target events" (i.e. events of interest like a failure that needs to be proactively handled).

As depicted in the Figure, Data Preprocessing consists of two sub-processes, the (i) feature extraction and (ii) the cleaning and integration processes. Considering the former process (i.e. feature extraction) both statistical and manual techniques can be applied to reduce directly the size (e.g. data sampling) or the dimensionality of the data. Factor Analysis (FA), Principal Component Analysis (PCA), Singular Value Decomposition (SVD), or Latent Semantic Analysis

(LSA) are widely used statistical techniques, which are able to reveal hidden combinatorial variables, lowering the complexity and the size of the data to be processed. Engineer experts' contribution is also important to pin point correlations between measurements of the same data source or dependencies between measurements and events of different collected datasets. Their knowledge can also assist the filtering process of data points that not contribute or complicate the analysis (i.e. noise) or the selection of specific features for the analysis. Cleaning and integration process, considers the handling of missing or incorrect entries, the scaling and normalization of the data and their transformation to compatible forms.

The data produced in the Data Preprocessing step are fed into the Analytical Processing step to initialize the data-driven techniques. There are several techniques that can be used, where each one maps the predictive maintenance problem to a wider family of problems with known solutions. For example, the vast majority of failure prediction methods try to solve a classification problem; given the events that occur within a time window, the classifier decides if a failure will occur within a predefined interval. Characteristic research works that use classification techniques are: [106] that utilizes a genetic algorithm approach, [107] where decision trees are proposed, [108] where Bayesian classification is used, [109], which utilizes Long-Short Term Memory (LSTM) neural networks and [110], where an approach based on frequent episode mining and Hidden Markov Models (HMMs) is proposed. Other research works use statistical methods to predict a failure, estimating the interval between the failures [111], or finding repeating suspicious patterns of events that preceded failures [112]. A recent work that applies predictive maintenance in aviation is [113], where authors use LU models for feature selection, Multiple Instance Learning (MIL) for training a Regression Model, and a Random Forest algorithm for solving the model. Their approach outperforms a widely used approach, which leverages linear Support Vector Machine (SVM) trained in the same MIL context. Another recent research study [104], combines data-driven modeling and Auto-Regressive Moving Average (ARMA) forecasting, to predict the next fault event based on past history events. ARMA forecasting, statistics raw features and PCA are used as an ensemble to train the data-driven model. They evaluate five popular machine learning techniques (i.e. k-Nearest Neighbors (kNN), Random Forests (RF), Neural Networks (NN), Support Vector Machines (SVM) and generalized Linear Regression (gLR)), which are compared to a LU model using Weibull distribution. The results show that SVM combined with ARMA forecasting can outperform traditional LU models.

The techniques presented depend on historical data, however there are also state-of-the-art techniques that can be applied on streaming data, like SPADE [114] for pattern mining in streaming time series, or MCOD [115] for outlier detection in data streams and MatrixProfile [116] for motif discovery in data series.

These techniques are going to be tested in the context of the BOOST 4.0 project, in order to produce more advanced predictive maintenance techniques capable of handling big data.

Know-How in Big Data Handling and appropriate technologies, as well as big data infrastructure is essential to the BOOST 4.0 project. Within Benteler, currently the following data storages are in place, collecting data relevant to smart maintenance

- Machine Data Lake with streaming data (e.g. temperatures, engine speeds, motor currents, forces, pressures, etc.) (Approx. 100 GB for five machines per month)
- Maintenance Database. Data Lake with logs and protocols of maintenance activities over 1 GByte –historical data of maintenance actions in pilot plant.
- Manufacturing Execution Data. Manufacturing data logs comprising order, parts and product information, cycle times and etc. over 90 GByte –historical manufacturing data for pilot plant

Furthermore, the following Big Data Analytics tools are state of the art and are suitable for application in the trial implementation:

- Hadoop, Spark
- Influx DB - Time series Data Base
- Grafana – Visualisation of Timeseries Data
- Python-based tools: Jupyter, Spyder, Anaconda distribution/environment. Libraries including SciPy, pandas, scikit-learn
- open communication tools (OPC-UA, International Data Space)

The trial technology providers are also experienced in the implementation of specialized data analytics solutions e.g. expert systems, virtual sensors, condition monitoring and process monitoring. Some examples are:

- expert system for industrial separators: detection of failure states based on analysis of vibration data with neural networks; recommendation of optimized process parameters based on formalized expert information; automatic process optimization using empiric cost-functions in the machine parameter space
- Condition monitoring for complex chemical processes: inline-measurement of viscosity of adhesives with a virtual sensor, based on machine data of the reactor (training data input) and laboratory results (training data output). Process modelling with neural networks.
- Optimization of quality control: data analysis of quality control data for maximizing time interval between consecutive quality control events
- Process automation of a dough kneader: chemical processes during dough kneading are modelled based on expert knowledge of experienced bakers (typical kneading time) and machine data using neural networks
- Process models and structuring tools for use case implementation, e.g. Analytics Layer Model, Data Analytics Canvas for use case design, Data Map for analysis of data sources

along business processes, process-x (-monitoring, -parametrization, -optimization) taxonomy; extended CRISP-DM cycle

11.2 Trial present scenario

The present maintenance scenario is different for scheduled maintenance routine and unplanned repairs. Process descriptions are given in graphical notation (OMEGA) in Figure 11-2, Figure 11-3 and Figure 11-4.

Maintenance Process: Scheduled Inspection / Maintenance

- prepare maintenance plan / schedule
- check supplies
- check/plan human resources availability
- planned production halt, if necessary
- do maintenance
- maintenance protocol
- hand over to production division

Maintenance Process: Unplanned repair

- failure notification / maintenance order
- production halt, if necessary
- find cause for failure
- check spare parts / supplies;
- order spare parts / supplies, if necessary
- do repair / maintenance
- repair / maintenance protocol
- hand over to production division

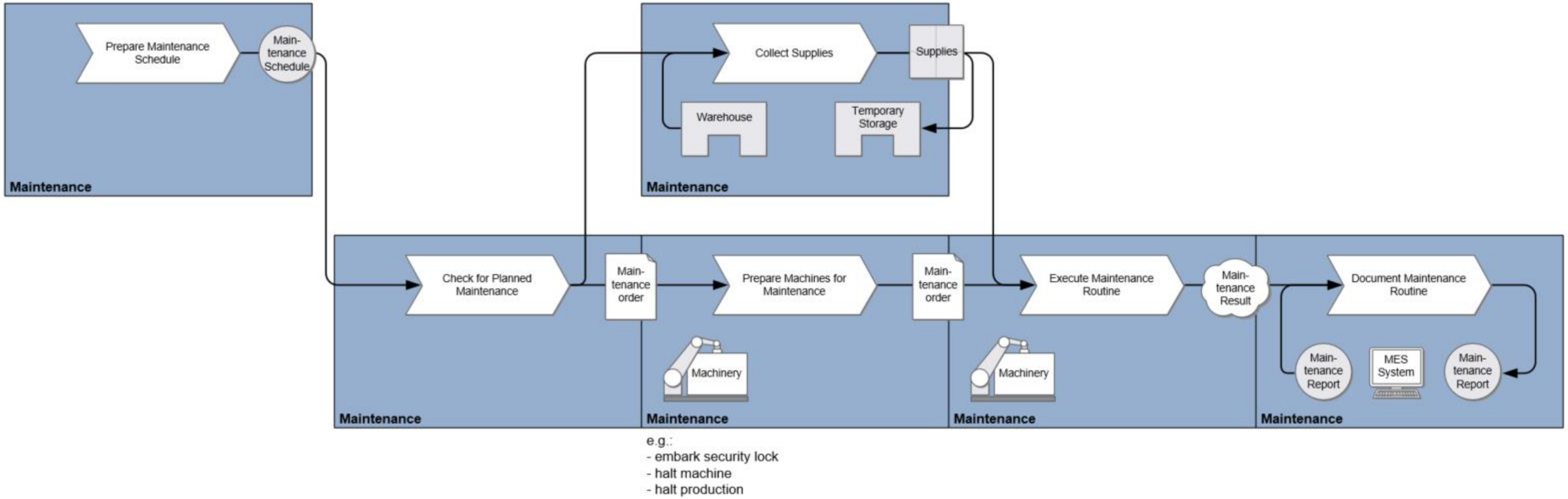


Figure 11-2 Maintenance Process for Scheduled Inspection

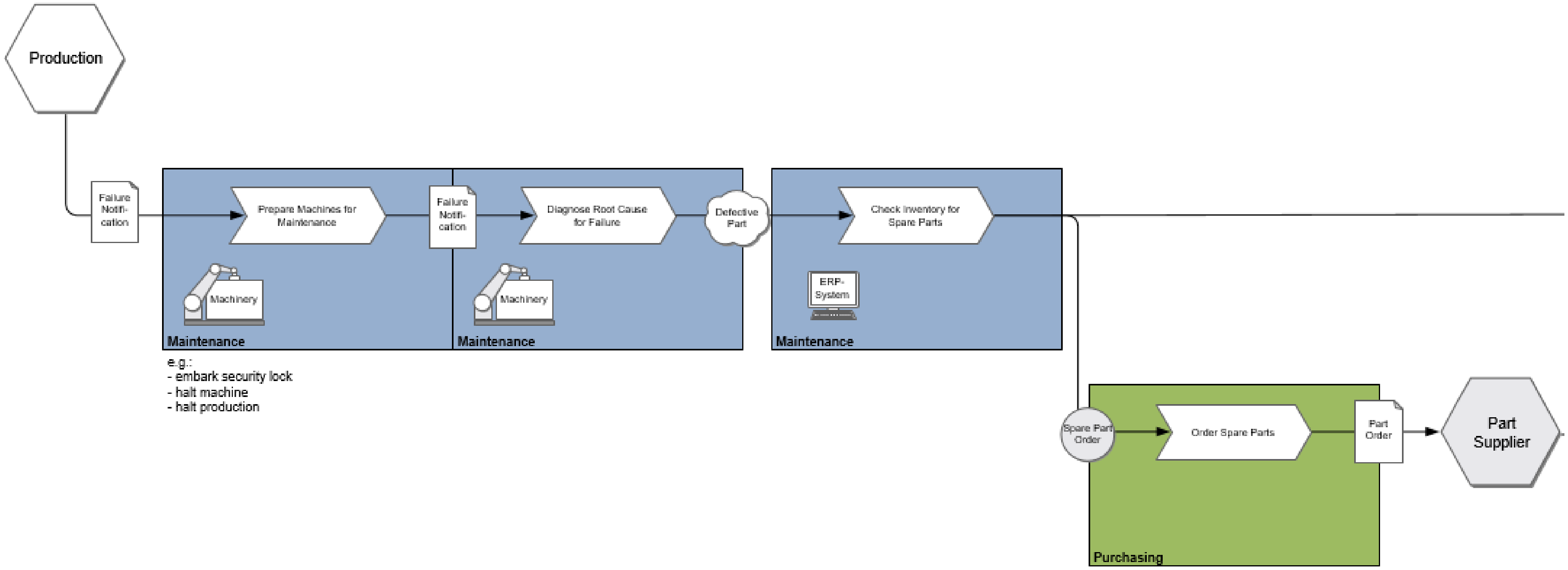


Figure 11-3 Maintenance Process for Unplanned Repair (1)

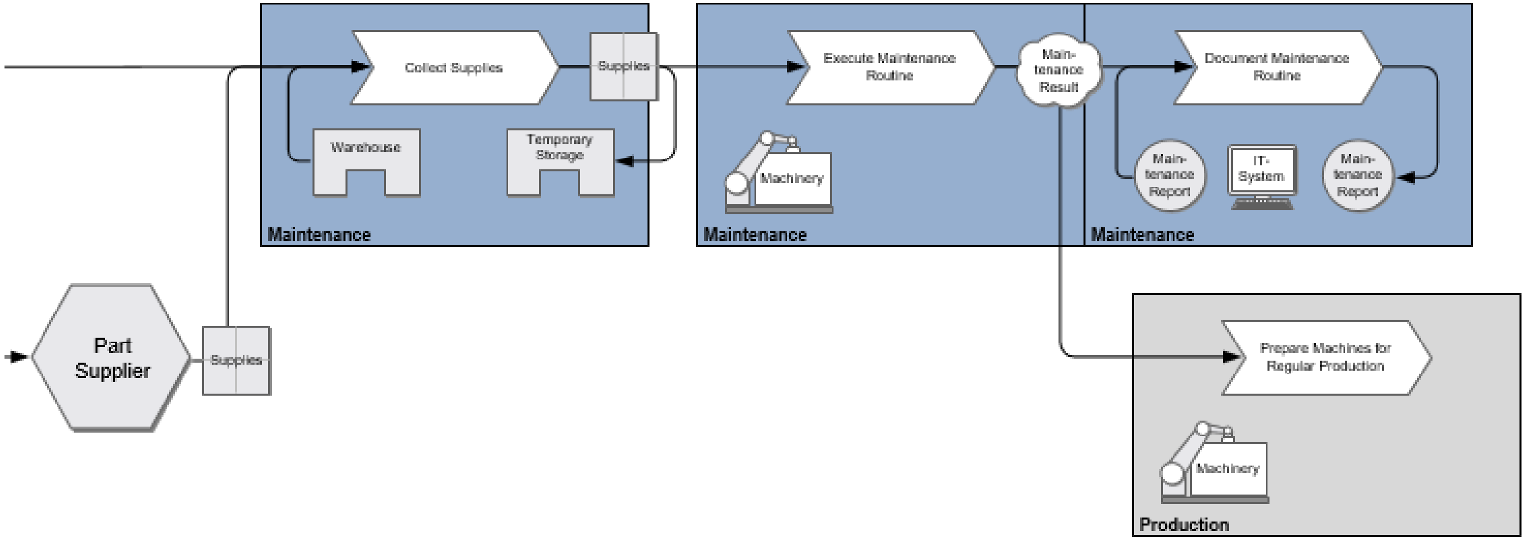


Figure 11-4 Maintenance Process for Unplanned Repair (2)

11.3 Weaknesses and bottlenecks

Weaknesses and bottlenecks may be linked with 5 "M"s in the case of the Benteler production line.

1. Manpower: Availability and suitability of human resources of the maintenance department.
2. Material: Spare parts, tools, consumables and other supplies.
3. Method: of production, of maintenance scheduling, of collaboration among the production and maintenance divisions.
4. Measurements: type and format of data coming from machines and personnel. Possibly not enough, not conclusive data with no records from maintenance
5. Machine: restrictions from the machine manufacturers as to the type/level of maintenance/repair that the machine themselves can undergo by typical maintenance personnel in automobile industry.

WEAKNESS & BOTTLENECKS	DESCRIPTION	AREA
Maintenance schedule not satisfied	Maintenance schedule cannot be satisfied due to additional unplanned maintenance events	Manufacturing
Low on supplies / Out of supplies	Consumable materials out of stock due to unplanned maintenance events	Manufacturing / Purchasing
Out of spare parts	Spare part not available in case of unplanned failure or break-down	Manufacturing / Purchasing
No maintenance resources	No maintenance staff available due to unplanned maintenance events	Manufacturing
Unplanned production halt	Unplanned production halt, delay in production / production not in time	Manufacturing / Sales
Maintenance/repair not properly documented		Manufacturing
Out of tools / Non-available tools	Necessary tools are not available to the maintenance personnel. / Tools may be	Manufacturing / Purchasing

	available at factory level, but they are used elsewhere	
No suitable maintenance resources	Maintenance personnel not trained enough to perform specific tasks. Mixture of maintenance teams fails to include experienced and novice workers.	Manufacturing / Management

11.4 Trial future scenario

- Prediction of machine break down --> supplies can be pre-ordered; resources can be planned; maintenance schedule can be kept
- Assistance for root cause diagnosis --> faster repair times, shorter production halt

11.5 Expected results

- Health monitor allows production to detect problems early before failures occur
- With the help of the framework even non-analytics-experts can add and analyze new issues, sensors etc.
- Transfer to other plants, machines etc.

11.6 Execution plan

The execution plan for the different business processes is shown in the following table:

Step	Actions	M1-M9	M10-M18	M19-M30	M31-M36
BUSINESS SCENARIO 1: Maintenance of a Press System					
BUSINESS PROCESS 1 - Maintenance in case of oil leakage					
1	Identify requirements and the most relevant data of a press to predict a maintenance task				
2	Develop, deploy and evaluate algorithm for anomaly detection on a press				
3	Develop, deploy and evaluate algorithm for prediction of a maintenance task				
4	Transfer, evaluation and KPIs collection				
BUSINESS PROCESS 2 – Maintenance on a scrap belt					
1	Identify requirements and the most relevant data of a scrap belt to predict a maintenance task				
2	Develop, deploy and evaluate algorithm for anomaly detection on a scrap belt				

3	Develop, deploy and evaluate algorithm for prediction of a maintenance task				
4	Transfer, evaluation and KPIs collection				
BUSINESS PROCESS 3 – Tech Provider					
1	Analysis of the Maintenance Processes and technical systems				
2	Modelling for fault detection & prediction and validation				
3	Model deployment				
4	Model transfer				
5	Evaluation and KPIs collection				
BUSINESS PROCESS 4 – Technology Center					
1	EIDS architecture design				
2	EIDS connectors implementation				
3	EIDS connectors functionality validation				
4	EIDS connectors – EIDS AppStore communication				
5	Evaluation and KPIs collection				

12 References

[1] D. Mihai, «Visual Components: and Advanced Production Line Simulation Software,» *Smashing Robotics,* 25-Jan-2016.

[2] F. Bonami, R. Nilito y J. e. a. Zhu, «Fog computing and its role in the internet of things,» *First edition of the MCC workshop on mobile cloud computing,* 2012.

[3] S. Goldstein, «Simple, open and standarised,» *PC-Control - The New Automation Technology Magazine,* nº 2, 2016.

[4] A. Pühringer, «How 'Industrial Cloud Communications' Delivers the Benefits of Internet-Connected Manufacturing,» *Whitepaper. Hilscher,* 2016.

[5] *The MathWorks, Inc., "OPC UA," 1994 - 2017. [Online]. Available: https://de.mathworks.com/discovery/opc-ua.html. [Accessed März 2017]..*

[6] *Schmidhuber, J. (2014). Deep learning in neural networks: An overview. Neural Networks, 85 - 117..*

[7] *Gruber, T. R. (1993). A translation approach to portable ontology specifcations. Knowl. Acquis., 5(2):199-220..*

[8] *Chandrasekaran, B., Josephson, J., and Benjamins, V. (1999). What are ontologies, and why do we need them? Intelligent Systems and their Applications, IEEE, 14(1):20-26..*

[9] *Gene Ontology Consortium. "The gene ontology project in 2008." Nucleic acids research 36.suppl 1 (2008): D440-D444..*

[10] *Available online: http://disease-ontology.org/ [Accessed März 2017]..*

[11] *Cheng, Chih-Hong, et al. "Semantic degrees for industrie 4.0 engineering: Deciding on the degree of semantic formalization to select appropriate technologies." Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering. ACM, 2015..*

[12] *Adolphs, P., Bedenbender, H., Dirzus, D., Ehlich, M., Epple, U., Hankel, M., Heidel, R., Hoffmeister, M., Huhle, H., Kärcher, B. and Koziolek, H., 2015. Reference architecture model industrie 4.0 (rami4. 0). VDI/VDE Society Measurement and Automatic Contr.*

[13] *Mahmood, Y. A., Ahmadi, A., Verma, A., Srividya, A., & Kumar, U. (2013). Fuzzy fault tree analysis: a review of concept and application. International Journal of System Assurance Engineering and Management, 19-32.*

[14] *Ontology Systems, "Rethinking Data Integration in a Post-Google World," 2013.*

[15] *Ontology Systems, "Ontology, Big Data and the Logical Data Warehouse," 2013..*

[16] *Ontology Systems, "Ontology's Common Cause Analysis (CCA) Module Datasheet", Oktober 2016. [Online]. Available: https://www.ontology.com/resources/data-sheets/?show_form=true&file_id=489. [Accessed März 2017]..*

[17] *T. Gavrilova and M. Gladkova, "Big Data Structuring: The Role of Visual Models and Ontologies," in Proceedings of the Second International Conference on Information Technology and Quantitative Management, ITQM 2014, Moscow, 2014..*

[18] *Obitko, Marek, and Václav Jirkovský. "Big data semantics in industry 4.0." International Conference on Industrial Applications of Holonic and Multi-Agent Systems. Springer International Publishing, 2015..*

[19] *Compton, M., et al. "The SSN ontology of the W3C semantic sensor network incubator group." Web semantics: science, services and agents on the World Wide Web 17 (2012): 25-32..*

[20] *Grangel-González, Irlán, et al. "Towards a semantic administrative shell for industry 4.0 components." Semantic Computing (ICSC), 2016 IEEE Tenth International Conference on. IEEE, 2016..*

[21] *Structure of the Administration Shell - Plattform Industrie 4.0, Federal Ministry for Economic Affairs and Energy (BMWi) Public Relations Germany, Available: https://www.plattform-i40.de/I40/Redaktion/EN/Downloads/Publikation/structure-of-the-administrati.*

[22] *Girardi, D., Küng, J., Kleiser, R., Sonnberger, M., Csillag, D., Trenkler, J., & Holzinger, A. (2016). Interactive knowledge discovery with the doctor-in-the-loop: a practical example of cerebral aneurysms research. Brain Informatics, 1-11..*

[23] *Wartner, S., Girardi, D., Wiesinger-Widi, M., Trenkler, J., Kleiser, R., & Holzinger, A. (2016, September). Ontology-Guided Principal Component Analysis: Reaching the Limits of the Doctor-in-the-Loop. In International Conference on Information Technology.*

[24] *Girardi, D., Kueng, J., and Holzinger, A. (2015). A Domain-expert Centered Process Model for Knowledge Discovery in Medical Research: Putting the Expert-in-the-Loop. In Brain Informatics and Health, pages 389-398. Springer International Publishing..*

[25] *Gorecky, Dominic, et al. "Human-machine-interaction in the industry 4.0 era." Industrial Informatics (INDIN), 2014 12th IEEE International Conference on. IEEE, 2014..*

[26] «Enterprise integration and interoperability in manufacturing systems: Trends and issues,» *Computers in Industry,* vol. 59, nº 7, pp. 641-646, 2008.

[27] G. Z. e. al, «Model-based approaches for interoperability of next generation enterprise information systems: state of the art and future challenges,» *Inf. Syst. E-Business Management,* vol. 15, nº 2, p. 2017.

[28] F. V. a. B. M. Giovanni Giachetti, «Interoperability for model-driven development: Current state and future challenges,» de *Sixth International Conference on Research Challenges in Information Science*, Valencia, 2012.

[29] E. N. e. al, «Requirements and languages for the semantic representation of manufacturing systems,» *Computers in Industry,* vol. 81, pp. 55-66, 2016.

[30] N. C. e. al, «A model-driven ontology approach for manufacturing system interoperability and knowledge sharing,» *Computers in Industry,* vol. 64, nº 4, pp. 392-401, 2013.

[31] M. C. e. al, «The SSN ontology of the W3C semantic sensor network incubator group,» *J. Web Sem,* vol. 17, pp. 25-32, 2012.

[32] M. D. a. A. T. Herve Panetto, «ONTO-PDM: Product-driven ONTOlogy for Product Data Management interoperability within manufacturing process environment,» *Advanced Engineering Informatics,* vol. 26, nº 2, pp. 334-348, 2012.

[33] P. P.-S. a. B. M. Bernardo Cuenca Grau, «OWL 2 Web Ontology Language Document Overview (Second Edition),» W3C, 2012.

[34] J. A. Hendler, «A New Portrait of the Semantic Web in Action,» *IEEE Intelligent Systems,* vol. 23, nº 3, 2008.

[35] D. L. P. e. al, «A middleware framework for scalable management of linked streams,» *J. Web Sem,* vol. 16, pp. 42-51, 2012.

[36] P. R. a. H. Paulheim, «Semantic Web in data mining and knowledge discovery: A comprehensive survey,» *J. Web Sem,* vol. 36, pp. 1-22, 2016.

[37] A. K. e. al, «Challenges of data integration and interoperability in big data,» de *2014 IEEE International Conference on Big Data*, Washington, DC, 2014.

[38] S. A. e. al, «The BigDataEurope Platform - Supporting the Variety Dimension of Big Data,» de *7th International Conference Web Engineering, ICWE 2017*, 2017, Rome.

[39] G. S. B. a. S. W. Julian Schutte, «Der Trusted Connector im Industrial Data Space,» 2018. [En línea]. Available: https://arxiv.org/abs/1804.09442.

[40] P. Russom, TDWI Best Practice Report: Big Data Analytics, TDWI Research, 2011.

[41] J. Shafer, R. Agrawal y M. Metha, «SPRINT: A Scalable Parallel Classifier for Data Mining,» de *Proceedings of 22th International Conference on Very Large Data Bases*, Mumbai, India, 1996.

[42] D. Luo, C. Ding y H. Huang, «Parallelization with Multiplicative Algorithms for Big Data Mining,» de *IEEE International Conference on Data Mining*, Brussels, Belgium, 2012.

[43] R. Chen, K. Sivakumar y H. Kargupta, «Collective Mining of Bayesian Networks from Distributed Heterogeneous Data,» *Knowledge and Information Systems,* vol. 6, nº 2, pp. 164-187, 2004.

[44] X. Wu, X. Zhu, G.-Q. Wu y W. Ding, «Data Mining with Big Data,» *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,* vol. 26, nº 1, pp. 97-107, 2014.

[45] C. T. Chu, S. K. Kim, Y. A. Lin, Y. Yu, G. R. Bradsky, A. Y. Ng y K. Olukotun, «Map-Reduce for Machine Learning on Multicore,» de *Proceedings of the 20th Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, 2006.

[46] C. Ranger, R. Raghuraman, A. Penmetsa, G. Bradski y C. Kozyrakis, «Evaluating MapReduce for Multi-Core and Multi-processor Systems,» de *Proceedings of the IEEE 13th International Symposium on High Performance Computer Architecture*, Phoenix, AR, USA, 2007.

[47] D. Gillick, A. Faria y J. DeNero, MapReduce: Distributed Computing for Machine Learning, Berkley, 2006.

[48] A. Ghoting y E. Pednault, «Hadoop-ML: An Infrastructure for the Rapid Implementation of Parallel Reusable Analytics,» de *Proceeding of the Large-Scale Machine Learning: Parallelism and Massive Data Sets Workshop (NIPS '09)*, 2009.

[49] H. Tong y U. Kang, «Big Data Clustering,» de *Data Clustering: Algorithms and Applications*, Boca Raton, FL, USA, CRC Press, 2013, pp. 259-274.

[50] A. S. Hashmi y T. Ahmad, «Big Data Mining: Tools & Algorithms,» *International Journal of Recent Contributions from Engineering, Science & IT,* vol. 4, nº 1, pp. 36-40, 2016.

[51] The Apache Software Foundation, «Apache Mahout,» The Apache Software Foundation, 2017. [En línea]. Available: https://mahout.apache.org/.

[52] The Apache Software Foundation, «Spark MLlib,» The Apache Software Foundation, 2017. [En línea]. Available: https://spark.apache.org/mllib/.

[53] P. Figueiras, R. Silva, A. Ramos, G. Guerreiro, R. Costa y R. Jardim-Gonçalves, «Big Data Processing and Storage Framework for ITS: A Case Study on Dynamic Tolling,» de *ASME 2016 International Mechanical Engineering Congress and Exposition*, 2016.

[54] G. Guerreiro, P. Figueiras, R. Costa , R. Silva y R. Jardim-Gonçalves, «Data Processing and Harmonization for Intelligent Transportation Systems: An Application Scenario on Highway Traffic Flows,» de *Learning Systems: From Theory to Practice*, Springer, 2018.

[55] P. Figueiras, G. Guerreiro, R. Costa, L. Bradeško, N. Stojanovic y R. Jardim-Gonçalves, «Big Data Harmonization for Intelligent Mobility: A Dynamic Toll-Charging Scenario,» de *OnTheMove EI2N 2016*, Rhodes, Greece, 2016.

[56] R. Costa, P. Figueiras, C. S. Gutierrez y L. Bradeško, «Personalized Intelligent Mobility Platform: An Enrichment Approach Using Social Media,» de *Intelligent Decision Technology Support in Practice*, Springer, 2016.

[57] R. Costa, P. Figueiras, G. Guerreiro, L. Bradeško, N. Stojanovic, P. Georgakis, E. Bothos y B. Magoutas, «Proactive recommendations for Intelligent Mobility - An approach based on real-time big data processing,» de *I-ESA 2016 - Interoperability for Enterprise Systems and Applications*, Guimarães, Portugal, 2016.

[58] C. S. Gutierrez, P. Figueiras, P. Oliveira, R. Costa y R. Jardim-Gonçalves, «An Approach for Detecting Traffic Events Using Social Media,» de *Emerging Trends and Advanced Technologies for Computational Intelligence*, 2016.

[59] R. Costa, P. Figueiras, R. Jardim-Gonçalves, J. Ramos-Filho y C. Lima, «Semantic enrichment of product data supported by machine learning techniques,» de *International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Madeira, Portugal, 2017.

[60] P. Figueiras, R. Costa, G. Guerreiro, H. Antunes, A. Rosa y R. Jardim-Gonçalves, «User interface support for a big ETL data processing pipeline an application scenario on highway toll charging models,» de *International Conference on Engineering, Technology and Innovation (ICE/ITMC)*, Madeira, Portugal, 2017.

[61] I. F. Vis, «Survey of research in the design and control of automated guided vehicle systems,» *European Journal of Operational Research,* vol. 170, nº 3, pp. 677-709., 2006.

[62] T. Müller, Automated Guided Vehicles, UK/Berlin.: IFS (Publications) Ltd./Springer-Verlag, 1983.

[63] L. Schulze, «Worldwide AGV-Systems of European Producers - Year-Analysis,» 2016.

[64] L. Schulze, S. Behling y L. Buhrs, «Automated guided vehicle systems: a driver for increased business performance,» de *International multiconference of engineers and computer scientists*, 2008.

[65] L. Xu, W. He y S. Li, «Internet of things in industries: A survey,» *IEEE Transactions on Industrial Informatics,* vol. 10, nº 4, pp. 2233-2243, 2014.

[66] F. Tao, Y. Zuo, L. Xu y L. Zhang, «IoT-Based intelligent perception and access of manufacturing resource toward cloud manufacturing,» *IEEE Transactions on Industrial Informatics,* vol. 10, nº 2, pp. 1547-1557, 2014.

[67] Y. Liu y X. Xu, «Industry 4.0 and cloud manufacturing: A comparative analysis,» de *ASME 11th International Manufacturing Science and Engineering Conference*, 2016.

[68] P. Ferrari, A. Flammini, E. Sisinni, S. Rinaldi, D. Brandão y M. S. Rocha, «Delay Estimation of Industrial IoT Applications Based on Messaging Protocols,» *IEEE Transactions on Instrumentation and Measurement,* vol. 99, pp. 1-12, 2018.

[69] M. Grieves, «Digital Twin: Manufacturing Excellence Through Virtual Factory Replication,» 2014. [En línea]. Available: http://innovate.fit.edu/plm/documents/doc_mgr/912/1411.0_Digital_Twin_White_Paper_Dr_Grieves.pdf.

[70] U. Margolin, A. Mozo, B. Ordozgoiti, B. Raz, E. Rosensweig y I. Segall, «Using Machine Learning to Detect Noisy Neighbors in 5G Networks,» 2016.

[71] B. Claise, «Specification of the IP flow information export (IPFIX) protocol for the exchange of IP traffic flow information (No. RFC 5101).,» *IETF.*

[72] A. Finamore, M. Mellia, M. Meo, M. M. Munafo, P. Di Torino y D. Rossi, « Experiences of internet traffic monitoring with tstat,» *IEEE Network,* vol. 25, nº 3, pp. 8-14, 2011.

[73] *M. Agyemang, K. Barker, R. Alhajj, "A comprehensive survey of numeric and symbolic outlier mining techniques", Intelligent Data Analysis 10 (6) (2006) 521-538..*

[74] *V. Chandola, A. Banerjee, V. Kumar, "Anomaly detection: A survey", ACM Computing Surveys, 41.*

[75] *V. Hodge, J. Austin, "A survey of outlier detection methodologies", Artificial Intelligence Review 22 (2) (2004) 85-126..*

[76] *F. Inglesias, T. Zseby, "Analysis of network traffic features for anomaly detection", Machine Learning 101 (2015) 59-84..*

[77] *M. Markou, S. Singh, "Novelty detection: a review-part 1: statistical approaches", Signal Processing 83 (12) (2003a), 2481-2497..*

[78] *A. Patcha, J.-M. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends", Computer Networks 51 (12) (2007) 3448-3470..*

[79] *Z. Bakar, R. Mohemad, A. Ahmad, M. Deris, "A comparative study for outlier detection techniques in data mining", Cybernetics and Intelligent Systems, IEEE Conference (2006) 1-6..*

[80] *M. Markou, S. Singh, "Novelty detection: a review-part 2: neural network based approaches", Signal Processing 83 (12) (2003b) 2499-2521..*

[81] *D. Savage, X. Zhang, P. Chou, Q. Wang, "Anomaly detection in online social networks", Social Networks 39 (2014) 62-70..*

[82] *C. O. Reilly, A. Gluhak, M. A. Imran, S. Rajasegarar, "Anomaly detection in Wireless Sensor Networks in a non-stationary environment", IEEE Communications surveys & Tutorials 16 (3) (2014) 1413-1432.*

[83] *Y. Zheng, S. Rajasegarar, C. Leckie, M. Palaniswami, "Smart car parking: temporal clustering and anomaly detection in urban car parking", 2014 IEEE Ninth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP),.*

[84] *T. Vafeiadis, E. Bora-Senta, D. Kugiumtzis, "Estimation of linear trend onset in time series", Simulation Modelling Practice and Theory, vol. 19, 2011, pp. 1384–1398..*

[85] *T. Vafeiadis, S. Krinidis, C. Ziogou, D. Ioannidis, S. Voutetakis, D. Tzovaras, "Robust malfunction diagnosis in process industry time series", Proceedings of IEEE 14th Conference of Industrial Informatics 2016;111-116..*

[86] *H. Alkhadafea, A. Al-Habaibehb, A. Lotfia, Condition monitoring of helical gears using automated selection of features and sensors, Volume 93, November 2016, Pages 164–177.*

[87] *F. Schnizler, T. Liebig, S. Mannor, G. Souto, S. Bothe, H. Stange, Heterogeneous Stream Processing for Disaster Detection and Alarming, 2014 IEEE International Conference on Big Data, pp. 914-923..*

[88] *A. Motamedi, A. Hammad, Y. Asen, Knowledge-assisted BIM-based visual analytics for failure root cause detection in facilities management, Automation in Construction vol. 43, 2014, pp. 73–83..*

[89] *Yao, J.T., Y.L. Li and C.L. Tan, |Forecasting the Exchange Rates of CHF vs USD Using Neural networks", Journal of Computational Intelligence in Finance, Vol.5, No.2., pp7-13, 1997..*

[90] *Y. Freund, R.E. Schapire, A desicion-theoretic generalization of on-line learning and an application to boosting, in: Computational Learning Theory, Springer, 1995, pp. 23–37..*

[91] *https://homes.cs.washington.edu/~marcotcr/blog/lime/.*

[92] *D. E. Goldberg, "Genetic algorithms in search, optimization and machine learning, Addison-Wesley Longman Publishing Co., Inc. Boston, MA, USA ©1989..*

[93] *LeCun, Yann. "LeNet-5, convolutional neural networks". Retrieved 16 November 2013..*

[94] *S. Guodong , S.J. Shin, S. Jain, Data analytics using simulation for smart manufacturing, In Proceedings of the 2014 Winter Simulation Conference, pp. 2192-2203. IEEE Press, 2014..*

[95] *B. Dean, Data in Motion: The Next Frontier For Manufacturers, available at http://goo.gl/RcFqXk.*

[96] *C. Weidig,P. Galambos, A.B. Csapó, P. Zentay, P.Z. Baranyi, Future Internet-based Collaboration in Factory Planning, Acta Polytechnica Hungarica 11(7) (2014) 157-177..*

[97] *M. Matzopoululos, Dynamic Process Modeling: Combining Models and Experimental Data to Solve Industrial Problems. In Process Systems Engineering, Wiley, Germany, 2010..*

[98] *B. Brad, M. Chui, J. Manyika, Are you ready for the era of 'big data, McKinsey Quarterly 4 (2011) 24-35.*

[99] *W. Ribarsky, D.X. Wang, W. Dou, W.J. Tolone, Towards a Visual Analytics Framework for Handling Complex Business Processes, In System Sciences (HICSS), 2014 47th Hawaii International Conference on, 1374-1383. IEEE, 2014..*

[100] *M. Abish, R. Maciejewski, T.F. Collins, D.S. Ebert, Visual analytics law enforcement toolkit, In Technologies for Homeland Security (HST), 2010 IEEE International Conference on, 222-228. IEEE, 2010..*

[101] *S.G. Eick, Geospatial visualization with VisTracks, Wiley Interdisciplinary Reviews: Computational Statistics 2 (3) (2010) 272-286..*

[102] M. Schwabacher, «A survey of data-driven prognostics,» de *Infotech@ Aerospace*, 2005.

[103] L. Chi-chao, A comparison between the Weibull and lognormal models used to analyse reliability data (PhD Thesis), Nottingham: University of Nottingham, 1997.

[104] M. Baptista, S. Sankararaman, P. I. de Medeiros, C. J. Nascimento, H. Prendinger y M. E. Henriques, «Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling,» de *Computers & Industrial Engineering*, 2018.

[105] C. C. Aggarwal, Data mining: the textbook, Springer, 2015.

[106] G. M. Weiss y H. Hirsh, «Learning to Predict Rare Events in Event Sequences,» de *KDD*, 1998.

[107] G. Xiaohui, S. Papadimitriou, S. Y. Philip y S. P. Chang, «Online failure forecast for fault-tolerant data stream processing,» de *ICDE*, 2008.

[108] L. Yu, Z. Zheng, Z. Lan y S. Coghlan, «Practical online failure prediction for blue gene/p: Period-based vs event-driven,» de *Dependable Systems and Networks Workshops (DSN-W)*, 2011.

[109] K. Zhang, J. Xu, M. R. Min, G. Jiang, K. Pelechrinis y H. Zhang, «Automated IT system failure prediction: A deep learning approach,» de *IEEE International Conference on Big Data*, 2016.

[110] S. Laxman, V. Tankasali y R. W. White, «Stream prediction using a generative model based on frequent episodes in event sequences,» de *SIGKDD*, 2008.

[111] Y. Yuan, S. Zhou, C. Sievenpiper, K. Mannar y Y. Zheng, «Event log modeling and analysis for system failure prediction,» de *IIE Transactions*, 2011.

[112] Y. Watanabe, H. Otsuka, M. Sonoda, S. Kikuchi y Y. Matsumoto, «Online failure prediction in cloud datacenters by real-time message pattern learning,» de *CloudCom*, 2012.

[113] P. Korvesis, S. Besseau y M. Vazirgiannis, «â€Predictive Maintenance in Aviation: Failure Prediction from Post-Flight Reports,» de *ICDE*, 2018.

[114] C. Yueguo, M. A. Nascimento, B. C. Ooi y A. K. Tung, «Spade: On shape-based pattern detection in streaming time series,» de *ICDE*, 2007.

[115] D. Georgiadis, M. Kontaki, A. Gounaris, A. N. Papadopoulos, K. Tsichlas y Y. Manolopoulos, «Continuous outlier detection in data streams: an extensible framework and state-of-the-art algorithms,» de *SIGMOD*, 2013.

[116] M. Linardi, Y. Zhu, T. Palpanas y E. J. Keogh, «Matrix profile X: Valmod-scalable discovery of variable-length motifs in data series,» de *SIGMOD*, 2018.

[117] FIWARE. [En línea].

[118] «ETSI NGSI-LD,» [En línea].

[119] «IDSA,» [En línea].

[120] «OASIS,» [En línea].

[121] «XAML,» [En línea].

[122] A. T. Schreiber y Y. Raimond, «RDF 1.1 Primer,» W3C Working Group Note, 2014.

[123] B. Otto, S. Lohmann, S. Steinbuß y A. Teuscher, «IDS Reference Architecture Model,» IDSA, Berlin, 2018.

[124] «Google,» [En línea]. Available: http://www.google.com.

[125] N. Srivastava, G. Hinton, A. Krizhevsky, Sutskever I. y R. Salakhutdinov, «Dropout: A simple way to prevent neural networks from overfitting,» *The Journal of Machine Learning Research,* 2014.

[126] D. Maltby, «Big Data Analytics,» de *74th Annual Meeting of the Association for Information Science and Technology (ASIST)*, New Orleans, USA, 2011.

[127] S. Srinivasa y S. Mehta, «Big Data Analytics,» de *hird International Conference, BDA 2014. Proceedings*, New Delhi, India, 2014.

[128] J. Zakir, T. Seymour y K. Berg, «Big Data Aalytics,» *Issues in Information Systems,* vol. 16, nº 2, pp. 81-90, 2015.

[129] P. F. Bertran, «SQL on Hadoop: A state of the art,» 1 April 2014. [En línea]. Available: http://www.datasalt.com/2014/04/sql-on-hadoop-state-of-the-art/.

[130] The Apache Software Foundation., "Hadoop," 2014. [Online]. [Accessed 2016].

[131] The Apache Software Foundation, "Apache Spark," 2014. [Online]. Available: https://spark.apache.org.

[132] The Apache Software Foundation, "Apache Storm," 2014. [Online]. Available: https://storm.apache.org/.

[133] MongoDB, Inc., "MongoDB," 2015. [Online]. Available: https://www.mongodb.org/.

[134] The Apache Software Foundation, "Apache Cassandra Project," 2015. [Online]. Available: http://cassandra.apache.org/.

[135] Apache Software Foundation, «Apache HBase,» Apache Software Foundation, 2007. [En línea]. Available: https://hbase.apache.org/.

[136] The Apache Software Foundation, “Apache Hive,” 2011. [Online]. Available: https://hive.apache.org/.

[137] *Z. Bakar, R. Mohemad, A. Ahmad, M. Deris, “A comparative study for outlier detection techniques in data mining”, Cybernetics and Intelligent Systems, IEEE Conference (2006) 1-6..*