Next-Generation Big Data-Driven Factory 4.0 Operations and Optimization: The Boost 4.0 Experience

Oscar Lázaro, Jesús Alonso, Philip Ohlsson, Bas Tijsma, Dominika Lekse, Bruno Volckaert, Sarah Kerkhove, Joachim Nielandt, Davide Masera, Gaetano Patrimia, Pietro Pittaro, Giuseppe Mulè, Edoardo Pellegrini, Daniel Köchling, Thanasis Naskos, Ifigeneia Metaxa, Salome Leßmann, and Sebastian von Enzberg

Abstract This chapter presents the advanced manufacturing processes and big data-driven algorithms and platforms leveraged by the Boost 4.0 big data lighthouse project that allows improved digital operations within increasingly automated and intelligent shopfloors. The chapter illustrates how three different companies have been able to implement three distinct, open, yet sovereign cross-factory data spaces

O. Lázaro (✉) · J. Alonso
Asociación de Empresas Tecnológicas Innovalia, Bilbao, Spain
e-mail: olazaro@innovalia.org; jalonso@innovalia.org

P. Ohlsson · B. Tijsma
Philips Consumer Lifestyle, Amsterdam, The Netherlands
e-mail: philip.ohlsson@philips.com; bas.tijsma@philips.com

D. Lekse
Philips Electronics Netherland, Eindhoven, The Netherlands
e-mail: dominika.lekse@philips.com

B. Volckaert · S. Kerkhove · J. Nielandt
Interuniversitair Microelectronica Centrum, Leuven, Belgium
e-mail: bruno.volckaert@ugent.be; sarah.kerkhove@ugent.be; joachim.nielandt@ugent.be

D. Masera
Centro Ricerche Fiat, Orbassano, Italy
e-mail: davide.masera@stellantis.com

G. Patrimia · P. Pittaro
Prima Industrie, Collegno, Italy
e-mail: gaetano.patrimia@primapower.com; pietro.pittaro@primapower.com

G. Mulè · E. Pellegrini
Siemens, Milan, Italy
e-mail: giuseppe.gm.mule@siemens.com; edoardo.pellegrini@siemens.com

E. Curry et al. (eds.), *Technologies and Applications for Big Data Value*,
https://doi.org/10.1007/978-3-030-78307-5_16

under a unified framework: the Boost 4.0 big data reference architecture and Digital Factory Alliance (DFA) service development framework.

Keywords Data Visualization · Data Quality · Data Lake · Big Data Stream Analytics · Brownfield · Trial · Predictive Maintenance 4.0 · Injection Moulding 4.0 · SUMA 4.0 · Intra-logistics

1 Introduction

The rapidly growing number of sensors, embedded systems and connected devices as well as the increasing horizontal and vertical networking of value chains result in a huge continuous data flow. In fact, *the manufacturing sector generates more data annually than any other sector in the EU or US economy, and the manufacturing industry (83%) expects data to have a big impact on decision-making in 5 years*. As highlighted by the European data strategy [1], by 2025, we will experience a 530% increase in global data volume from 33 zettabytes in 2018 to 175 zettabytes, and data will represent an economic value of 829 million € in the EU27 economy compared to the €301 million that it represented in 2018 (2.4% of the EU GDP).

Big Data will have a profound economic and societal impact in the Industry 4.0 sector, which is one of the most active industries in the world, contributing to approximately 15% of EU GDP. According to the World Economic Forum report on Digital Transformation of Industry [2], *Big Data is expected to take off in the consumer market to a value at stake of over $600 billion for industry and $2.8 trillion for society in improved customer service and retailing experience*. Moreover, the total value that companies can create in five key areas of data sharing is estimated to be more than $100 billion. In fact, 72% of the factories consider that sharing data with other manufacturers can improve operations and 47% find enhanced asset optimization to be the most relevant application area. Big Data as part of European Industrial Digitization could see manufacturing industry *add gross value worth 1.25 T€*—or more importantly *suffer the loss of 605 BEUR* in foregone

D. Köchling
BENTELER Automobiltechnik, Paderborn, Germany
e-mail: daniel.koechling@benteler.com

T. Naskos · I. Metaxa
ATLANTIS Engineering, Athens, Greece
e-mail: naskos@abe.gr; metaxa@abe.gr

S. Leßmann
it's OWL Clustermanagement, Paderborn, Germany
e-mail: s.lessmann@its-owl.de

S. von Enzberg
Fraunhofer Institute for Mechatronic Systems Design, Paderborn, Germany
e-mail: sebastian.von.enzberg@iem.fraunhofer.de

value added if it fails to incorporate new data, connectivity, automation and digital customer interface enablers and getting their digital manufacturing processes ready, i.e. cognitive and predictive, in automotive engineering and logistics. The European Commission foresees that *advanced analytics in predictive maintenance systems only could reduce equipment downtime by 50% and increase production by 20%.* Overall, *only the top 100 European manufacturers could save around 160 BEUR* thanks to improved error-correcting systems and the ability to adjust production in real time. Additionally, *10% production efficiency improvement* can be realized in top 100 EU manufacturers with an associated *265 BEUR* gain for the industry.

Despite the big data promises, interestingly (1) only 3% of useful manufacturing data is tagged and even less is analysed, (2) manufacturing industry are currently losing up to 99% of the data value they capture since evidence cannot be presented at the speed decisions are made and (3) only half of industry is currently using any data to drive decisions with a much lower 15% of EU industry employing Big Data solutions as part of value creation and business processes.

Boost 4.0 [3] is a European lighthouse initiative for the large-scale trial of big data-driven factories. The Boost 4.0-enabled Connected Smart Factory 4.0 vision is one where digital design technologies enable short times to market, resources are optimally planned, downtime is predicted and prevented, waste and defects are eliminated, surplus production is minimized, machine behaviour is optimized as conditions change and systems can make context-based 'next best' actions. Connected devices in the factory report their status, giving operations personnel and decision-makers access to real-time, actionable information. Wearable technology tracks employee location and status in case of emergency. A global ecosystem of partners ensures that specific parts are replenished based on automated, real-time needs analysis. *Data is at the heart of Industry 4.0, the experience economy and the manufacturing digital transformation towards 'servitised' product service systems and outcome-based digital business models; as opposed to traditional product ownership business models*. But the massively growing information flow brings little value without the right analytics techniques.

Full adoption of a data-driven Factory 4.0 has been largely hampered by: unclear ownership and access right definition in the data value chain; need to harmonise cross-border heterogeneous flows of data; limited availability of open datasets to feed industrial ecosystems; insufficient diffusion of advanced technologies to preserve data confidentiality and privacy. All these issues, which broadly relate to data sovereignty, remain largely un-addressed challenges by current Digital Manufacturing Platforms solutions. The lack of such reference framework has the following drawbacks:

- Manufacturing big data sets are highly heterogeneous in nature and are spread across the product and factory lifecycles.
- Manufacturing big data is highly unstructured, hard to analyse and distributed across various sectors and different stakeholders involved in the product and factory lifecycles.

- Data is usually duplicated across many digital manufacturing platforms and systems (data multi-homing), thereby making it difficult to maintain 'quality' and updated data to base decisions upon.
- Data analysis necessarily implies a loose of control over the use of data from the data owner since the transfer of data across digital platforms and enterprises is mandatory for data consolidation and processing.
- Often valuable data is measured for real-time use in specialized systems, but not stored for later processing or recorded in a way suitable for data collation across individual systems.
- Data-driven decision support is slow and contextualization of information is cumbersome and involves intensive manual operation on data sources.
- Data transactions (grant of data access rights, data transfer) are slow, mediated and cumbersome.
- Machine- and shopfloor-generated data is usually not ready for sharing with external stakeholders.
- Engineering, production, IT and IoT data remain as isolated silos that make costly and complex the development of smart services on top of smart products.

In addition, Big Data will be exponentially created, processed and stored in the coming years—see EU Data strategy projections above—but no single infrastructure, let alone a single stakeholder, can do the job on its own. Boost 4.0 has addressed the lack of European and global standards and an Industry 4.0 big data reference framework that ensures data sovereignty, while enabling the agile and value-driven creation of ad hoc trusted data networks across currently isolated consumer experience data, usage-context data, production and engineering data 'clouds' (Fig. 1).

1.1 Big Data-Centric Factory 4.0 Operations

Data-centric operations is one of the fundamental cornerstones of modern industrial automation technologies and also one of the bases for decision-making and control operation. While the use of statistical data analysis for control is well established, recently the diffusion of big data methodologies added a new dimension, by providing additional approaches to data-centric automation. A big data-centric approach to Factory 4.0 operations opens the door to migration from an asset-centric decision process towards truly real-time, predictive and coordinated multi-level process centric decision processes. Such *process-centric, holistic and integrated data space* for Factory 4.0 operations calls for significant improvements in speed, flexibility, quality and efficiency (Fig. 2).

Within the Boost 4.0 project, pilots have leveraged Industrial Internet of Things (IIoT), big data space technologies, advanced visualization, predictive analytics and collaborative AI engineering and decision support systems for the benefit of significant improved operations. As shown in Fig. 3, Boost 4.0 is consider-

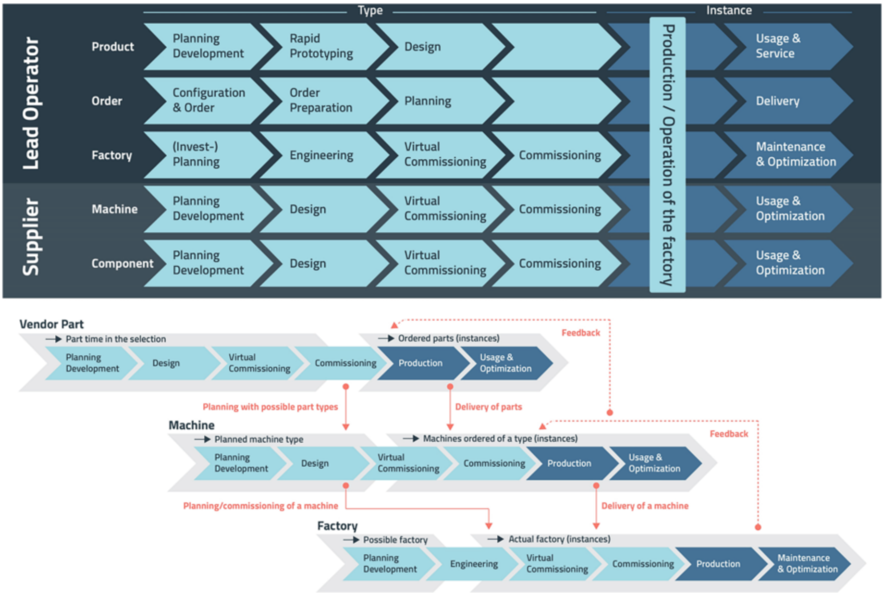

Fig. 1 Boost 4.0 'whole' lifecycle digital thread synchronization big data challenge

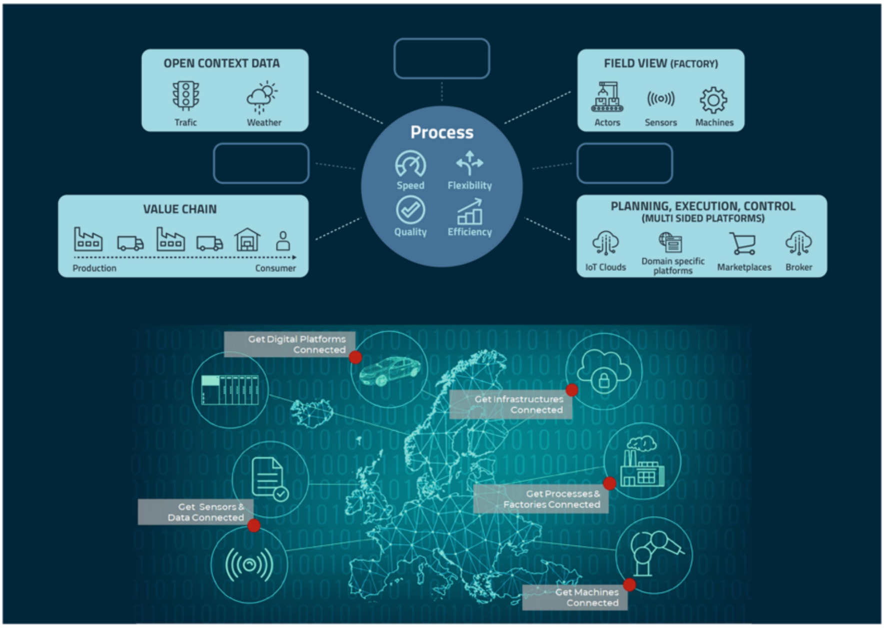

Fig. 2 Boost 4.0 'process-centric' data space connecting industrial things and platforms

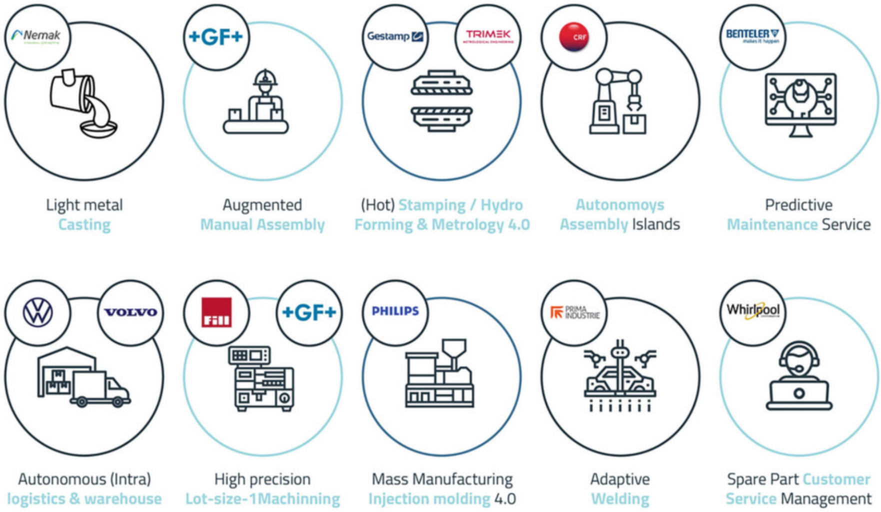

Fig. 3 Boost 4.0 manufacturing 4.0 processes supported

ing the development and evaluation of process-centric, data-centric, AI-powered advanced manufacturing 4.0 processes. Boost 4.0 is considering a highly diverse set of manufacturing 4.0 processes under a unified big data framework, ensuring high portability and replicability. The manufacturing 4.0 processes supported by Boost 4.0 range from light metal casting to augmented manual assembly, hot stamping, metrology 4.0, hydroforming, autonomous automated assembly islands, predictive maintenance, autonomous intra logistics and business network tracing, high-precision lot-size machining, mass manufacturing injection moulding 4.0, adaptive welding and spare part management customer services. Moreover, these processes are implemented across a number of sectors (automotive, white goods, high-end textiles, machine tool industry, ceramics, elevation, aero), thereby ensuring that highly varied sectors are amenable to big data transformations. The interested reader is referred to the additional chapters in this book for more details on the manufacturing processes implemented.

These Boost 4.0 data-driven manufacturing processes are supported by advanced big data technologies—e.g. data streaming, batch and predictive analytics, Machine Learning (ML) and Artificial Intelligence (AI)—which are applied seamlessly across the full product and process lifecycle (Smart Digital Engineering, Smart Digital Planning & Commissioning, Smart Digital Workplace & Operations, Smart Connected Production, Smart Service & Maintenance). Boost 4.0 has thereby leveraged a number of high-performance big data algorithms and platform features that, as illustrated by the implemented trials, can deliver high impact and performance improvements in factory operations; see Fig. 4.

The three Boost 4.0 lighthouse pilots that are presented in this chapter have introduced into their processes new big data methodologies to optimize different aspects

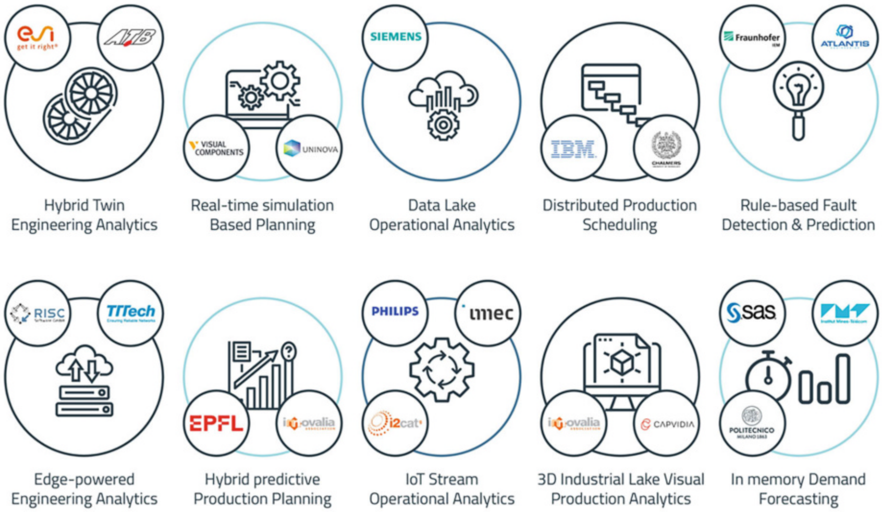

Fig. 4 Boost 4.0 big data algorithms and platforms

of the product lifecycle, from the production itself to the distribution of spares for the after-sales services. In the next section, Philips present their *Injection Moulding Smart Operations & Digital Workspace*, in which the Drachten premises pave the way for a generic platform usable for the full fleet of injection-moulding machines across Philips' factories. Afterwards, the BENTELER Automotive lighthouse trial is discussed, which deployed a big data platform for smart maintenance of industrial assets, focusing on the example of a hydraulic press. A novel predicted structured and effective approach toward assets' failure management and synchronization with higher level plant management system has been provided, where predicted failures and estimated RUL are dynamically assessed in real time for their severity and potential impact on the plant, evaluating their criticality in order to provide the right recommendation for remedy actions. Next, the FCA trial is introduced. In this use case, the focus is on the smart collaboration between mobile robots, more specifically AGVs and laser machine. Finally, some conclusions are drawn.

This chapter relates mainly to the technical priorities Data Analytics and Advanced Visualization and User Experience of the European Big Data Value Strategic Research & Innovation Agenda [4]. It addresses the horizontal concerns of analytics frameworks and processing, predictive and prescriptive analytics and interactive visual analytics of multiple-scale data of the BDV Technical Reference Model. It addresses the vertical concerns regarding the use of standards to facilitate integration of data end-points from legacy and heterogenous systems and development of trusted and sovereign data spaces across production sites and development of third-party applications and services. The work in this chapter relates mainly, but not only, to the Reasoning and Decision Making cross-sectorial technology enablers of the AI, Data and Robotics Strategic Research, Innovation & Deployment Agenda [5].

2 Mass Injection Moulding 4.0 Smart Digital Operations: The Philips Trial

2.1 Data-Driven Digital Shopfloor Automation Process Challenges

Philips Drachten encompasses a large suite of highly automated processes used during the manufacturing of electric shavers. Of these manufacturing processes, injection moulding is of particular importance, as it is used during the fabrication of plastic components for electric shavers. Injection moulding is a competitive market, which makes it essential for Philips Drachten to continuously improve on quality, production performance, and costs where this process is concerned.

All the plastic parts are manufactured on-site at Drachten, requiring approximately 80–90 moulding machines of multiple vendors, models and generations. For large manufacturing sites, generalization is key to deploy data-driven solutions. It is simply not feasible to develop a specialized solution for each machine in the machine park. Thus, in this pilot our main challenge is to develop scalable solutions.

Furthermore, specialized custom solutions do not yield a positive business case in the case of moulding; plastic is relatively cheap, meaning that fall-off is not that expensive. Building a solution per machine type would simply be too costly: the time investment required to build these custom solutions is too high compared to the potential annual savings. However, focusing on the fall-off rate of the entire plastic-part-making departments and all such departments, the financial gains are significant. By lowering the amount of time required to enable analytic capabilities for each machine, we can transform it into a positive business case. This is why in this pilot the focus has also been on developing general predictive maintenance and process control solutions that are cloud-enabled and thus easily scalable.

Another challenge that has been tackled is the interaction of data-driven digital processes with the current manufacturing processes and how data-driven decision should be translated into actionable insights within production.

From a strategic standpoint, it is expected that the technologies developed using data-driven processes can be developed into new autonomous modes of manufacturing. Production customization has been made possible, implying frequent product changeover and smaller batch sizes, so-called Innovative Big Data Cognitive Manufacturing Processes. This pilot has deployed a series of technologies that facilitate increased quality and productivity, while also investigating generalization and scalability of these technologies in an industrial setting.

2.2 Big Data-Driven Shopfloor Automation Value for Injection Moulding 4.0

Philips Drachten wants to remain a pilot location for Industry 4.0-related activities. The business experiences need to become more data driven, while the effort of achieving this should reduce over time. Over the years, there have been data-driven solutions demonstrated and implemented within production; however, they have never been successful in scaling up nor maintaining those data-driven solutions, as they have focused on special solutions for unique cases.

The data business process value lies in developing generic automated solutions capable of scaling up across multiple injection moulding machines. A failure prediction model is one example of a generic automated solution, which can be applied for multiple machines and it results in reduction of fall-off rate and reduction of machines' downtime.

The application of Big Data and fact-based decision-making, along with seamless connectivity in the manufacturing process, results in efficient ramp-up times between different moulds, along with full traceability along the process chain all the way to the customer. A new data collection and storage infrastructure has been deployed to effectively integrate various types of data into a single common repository. This includes state-of-the-art technologies like streaming, edge computing and cloud computing in order to provide our operators with actionable insights. The results of the data monitoring and machine learning must be made available to process engineers, assembly line operators and data scientists.

2.3 Implementation of Big Data-Driven Quality Automation Solutions for Injection Moulding 4.0

To successfully implement Big Data solutions within the production process, an architecture map was made for the pilot phase during Boost 4.0. From this pilot setup, we identified the basic components and tested the concepts of connecting machines to a 'data collection' platform. In addition, several technical elements were identified that needed to be taken care of in order to build a fully scalable platform for Philips' injection moulding machines.

Although the eventual end goal is to prepare for a generic platform usable for the full fleet of injection moulding machines across Philips' factories, the final architecture of this trial has been instantiated for the Drachten site only. This also allows building an on-premise platform to manage all local data and consider offloading and/or management connection to external platforms, e.g., the cloud and/or the Industrial Data Spaces framework (IDS). The Drachten facility is a so-called brown-field factory, which means we need to comply with the local architecture as implemented. Philips has teamed up with two technology providers, Philips Research (PEN) and IMEC, to support the implementation. PEN has a long heritage of pioneering innovation (inventions related to x-ray, optical recording, CD, DVD, etc.), currently focusing on data-driven research and service orientation, and IMEC is a world-leading research and innovation hub in nanoelectronics and digital technologies, combining widely acclaimed leadership in microchip technology and profound software and ICT expertise. Both technical partners (PEN & IMEC) provided input on where to put their proposed solutions:

- **Cloud connectivity** using a custom-build gateway service (based on the Microsoft Edge framework).
- **Data broker** to allow easy data acquisition for (historical) data, used for data analysis and machine learning models.
- **Machine learning models** that use real-time and historical data for predicting failures of machines.
- **Dashboard** (real-time) visualization of machine data, including pre-processing and machine learning models deployed as services.

Boost 4.0 big data platforms and techniques (Fig. 6) comply with RAMI 4.0 Digital Factory Alliance (DFA) service development reference architecture and ISO 20547 Big Data Reference Architecture (see chapter 'Big Data Driven Industry 4.0 Service Engineering Large Scale Trials: The Boost 4.0 Experience' in this book). Thus, it is possible to map into the Boost 4.0 Big Data Reference Architecture (RA) [6] the Philips predictive quality architecture (Fig. 5).

The Boost 4.0 architecture is based on multiple (Big Data) IT solutions being integrated via open APIs (Fig. 6). Some of which are essential and are part of the backbone, while others are optional and extend functionality beyond the core functionalities of the platform.

- **Machine connector**: Allow to acquire (time-series) data from the machine controller. This is highly dependent on the machine interfaces available on the equipment itself. Typical machine connectors include OPC(-UA), CodeSys, Serial, Modbus, Canbus, EtherCAT, etc.
- **Protocol translation:** Translate an industrial protocol to another (open) standard. In this case, this is done by KEPWARE [7] and transforms data to OPC-UA-formatted data.
- **Semantics injection:** Make data understandable by adding semantic information (standard names, units, location, source, etc.)
- **Streaming data ingestion:** Transform OPC-UA to JSON formatted data and put them on a Kafka bus.
- **Streaming data bus:** A publish/subscribe enabled pipeline for real-time data transport. In this case, the data is JSON, based on the JSON-Header-Body (JHB) standard for industrial applications.
- **Micro-service architecture** for deployment of (Docker) containers (connected to the data bus), usable for data (pre-)processing, data analysis and data visualization.

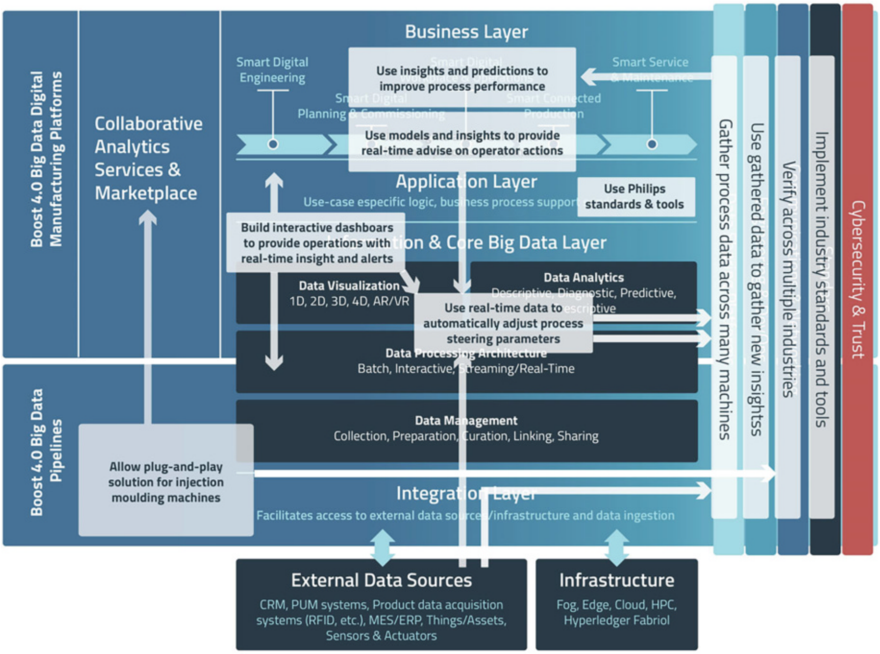

Fig. 5 Shopfloor automation trial mapping in Boost 4.0 big data reference architecture

- **A data historian** for long-term storage of time-series data. In the current version this is handled by Inmation [8] (based on MongoDB) or by Azure Time Series Insights (cloud storage, based on compressed JSON files in Parquet format).
- **Data broker** for providing different users with data from different sources in a standardized format, used for analysis. It supports real-time connections and is custom built.
- **Data analysis** is mainly taken care of by Python code (deployed in a container). Depending on the solution, multiple packages are used (like Pandas, SciKit, Keras, TensorFlow, etc.).
- **Rancher** solution to manage all micro-service containers from an easy-to-use web interface.
- **Open-source tools** for visualization of (live) data, based on Web technology, including Vue.JS, Quasar Framework, HTML, etc.).

2.4 Big Data Shopfloor Quality Automation Large-Scale Trial Performance Results

The instantiation and deployment of the Boost 4.0 reference model and implementation of big data pipelines into the Philips Drachten factory has translated in

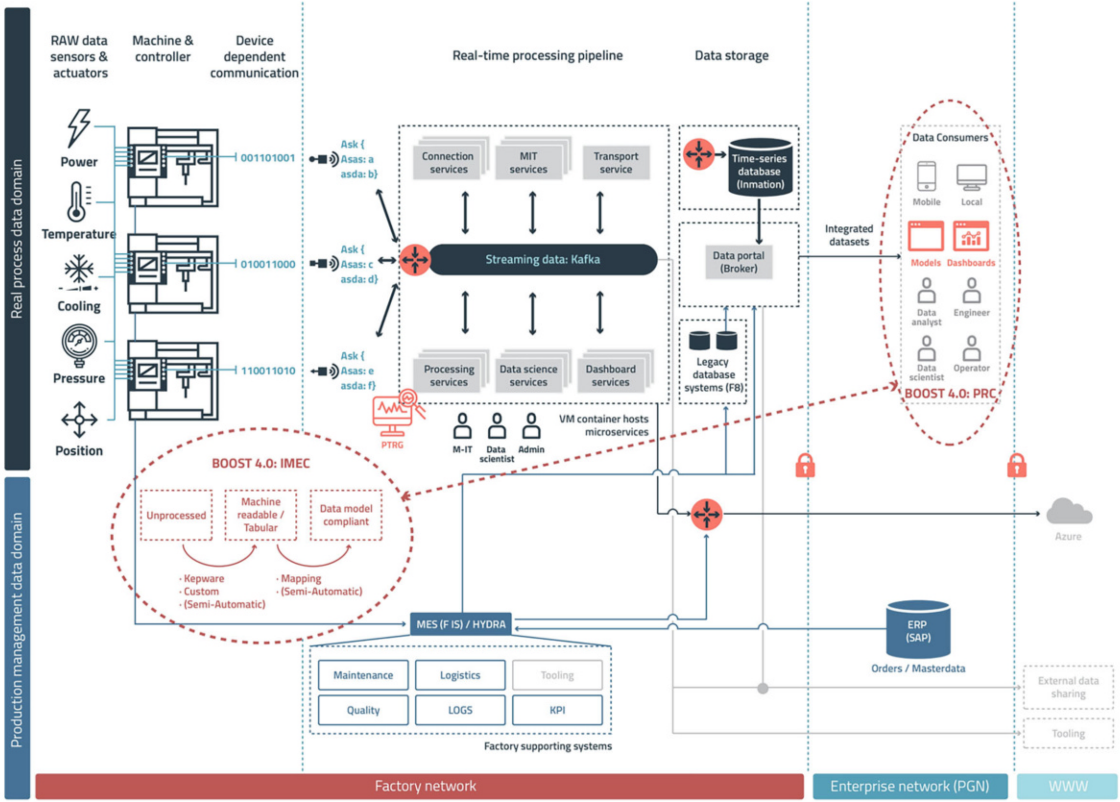

Fig. 6 Philips reference architecture and pipelines for predictive big data-driven quality automation solutions for injection moulding 4.0

Fig. 7 Shopfloor dashboards experience and advanced data processing tools

significant shopfloor performance improvements in terms of flexibility, efficiency, quality and time to market. This is directly related to the ability to implement advanced decision support dashboards that reduce decision-making time and allow anticipating unplanned events, leveraging close to 10% production performance improvements (Fig. 7).

The main quantitative efficiency achievements are summarized as follows:

- A 10% reduction in fall-off rate and a 9% reduction in downtime
- Increased availability of process parameters data available from every 20 minutes to real-time information and increased number of parameters from 10 to ~80 per machine per cycle
- Collected over 400k of individual shots in 5 months of injection moulding data, which can be used to build more advanced models
- By automating the process of machine data end-point, increased and more homogeneous data quality and decreased time needed for connecting a machine to the ‘real-time’ platform from 2 weeks to around 4 h
- Reduction of 70% in the amount of (non)valuable and unnecessary control actions by operators

The adoption of the Boost 4.0 universal big data pipeline has also translated into increased quality of work in the shopfloor in the following way:

- Providing technicians with a more efficient tool for solving production issues and becoming part of a *better method for troubleshooting*.
- Take *wiser and more informed decisions* based on facts, for instance avoiding acquiring new machines by using the current machine park more efficiently.
- Better understanding of the current state of the art regarding IT, semantics and machine Learning.

- Better understanding of the time-related behaviour of the injection moulding process.
- Use the pilot setup to showcase the value of digitalization to the management board of Philips Corporate, in order to obtain their attention and support.

2.5 Observations and Lessons Learned

In the first year of the project, some basic interfaces were deployed on the shopfloor. This provided valuable lessons on the exact requirements for deploying predictive quality processes on the shopfloor.

As the main goal is to provide the operators with valuable insights, it became clear that technology (IT) is only one part of the challenge. Working together with operators and productions engineers quickly results in other challenges. With the help of our partners, the technical implementation was built and deployed fast. It is of crucial importance to keep ‘operations’ in the loop at all times.

Bringing IT solutions to the shopfloor (and essentially making them part of the production system) also implies requirements that were not as visible at the start of the project. These requirements must make sure operation can rely on the performance as well as the availability of solutions, and include actions such as training, coaching, providing support, but also continuously monitor solutions, preferably 24/7. When these measurements are taken into consideration, the results of the experimentations will already have a significant impact on the quality control process.

3 Production Data Platform Trials for Intelligent Maintenance at BENTELER Automotive

BENTELER is a global, family-owned company serving customers in automotive technology, the energy sector and mechanical engineering. As an innovative partner, it designs, produces and distributes safety-relevant products, systems and services. In the 2019 financial year, Group revenues were €7.713 billion. Under the management of the strategic holding BENTELER International AG, headquartered in Salzburg, Austria, the Group is organized into the divisions BENTELER Automotive and BENTELER Steel/Tube. Around 30,000 employees at 100 locations in 28 countries offer first-class manufacturing and distribution competence—all dedicated to delivering a first-class service wherever their customers need it. BENTELER Automotive is the development partner for the world’s leading automobile manufacturers. Around 26,000 employees and more than 70 plants in about 25 countries develop tailored solutions for their customers. BENTELER Automotive’s products include components and modules in the areas of chassis, body, engine and exhaust systems, as well as solutions for electric vehicles.

3.1 Data-Driven Digital Shopfloor Maintenance Process Challenges

Intelligent, self-regulated maintenance is a key element in Industry 4.0. The networking of machines and plants and the availability of machine data allows continuous monitoring and evaluation of the health status of a production system in real time. Failures and malfunctions can be detected or even foreseen at an early stage, and measures to protect the functionality and performance of the production system can be derived from them. The aim of Smart Maintenance is to increase the performance of production technology, for example through increased plant availability, optimized process quality and improved planning.

The basic technologies for Smart Maintenance solutions are already available. Seventy per cent of machine and plant manufacturers are developing or piloting Smart Maintenance offerings or already offer them [9]. Market-ready solutions are offered in particular by component suppliers from the automation and drive technology sector, as they can be transferred to a large quantity of systems. Nevertheless, the application of Smart Maintenance in manufacturing is below expectations, even though solutions for individual components are available: On the side of machine operators, maintenance knowledge is required for a large number of different machine types and systems. This know-how is hardly ever bundled, documented or made available by means of standardized processes. According to Acatech [10], 47% of German manufacturing companies record information on malfunctions and failures only manually. Fifty-seven per cent of companies still initiate measures without any data at all. Only 4% make decisions based on real-time data.

The biggest challenge in developing a fault detection system is the availability of fault data. Compared to the total amount of data available, failures and errors occur only rarely. Many machine learning methods (especially so-called supervised learning methods) are therefore not or only partially applicable. Hence, mainly methods of anomaly detection were used during the development. Thereby characteristics for normal behaviour are derived from the signal courses in regular production use. During operation, deviations from this normal behaviour are detected and reported.

3.2 Implementation of a Big Data Production Platform for Intelligent Maintenance in Automotive

The goal of the trial is the implementation of a global (cross-factory) system for automated detection of failures and recommendations for actions in the context of machine health monitors with notification and planning of actions (Fig. 8). The availability of a platform for the storage and processing of production data is a prerequisite for the implementation of such centralized intelligent maintenance in

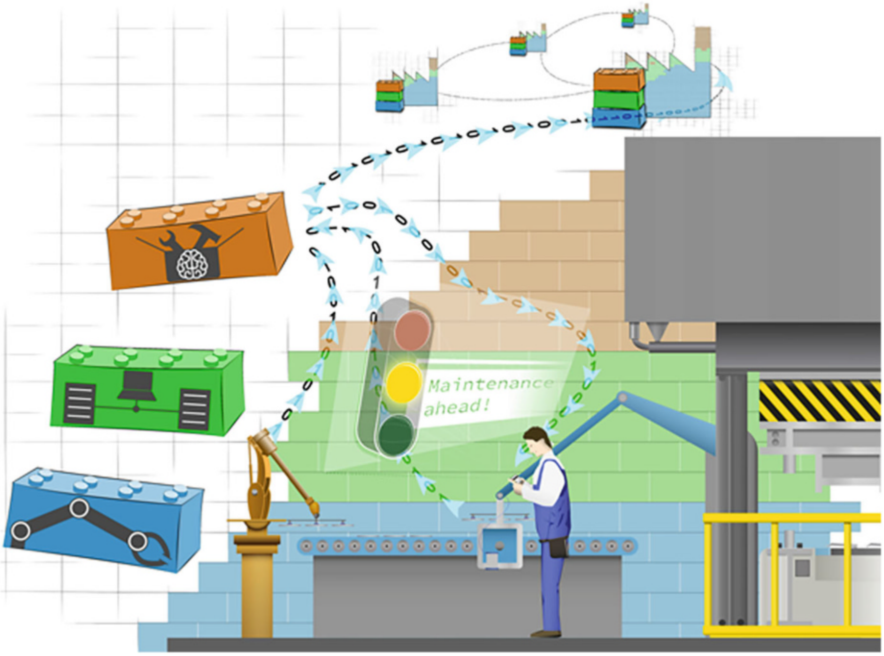

Fig. 8 Smart Maintenance Trial Factory: Solution modules from the systematic data connection, data infrastructure and intelligent data processing are the basis for the successful implementation of use cases (source: IEM)

production. As part of the Boost 4.0 project, BENTELER Automotive, the Fraunhofer Institute for Mechatronic Systems Design IEM and ATLANTIS Engineering have built a Smart Maintenance pilot factory within the Leading-Edge Cluster it's OWL—Intelligent Technical Systems OstWestfalenLippe.

Solutions for different problem dimensions have been developed. In addition to the technical infrastructure for industrial data analysis, the development of data evaluation, process integration on the application level and the methodical procedure for the implementation of Smart Maintenance have been analysed. The maintenance of a hydraulic press, as well as a material handling system have been considered as examples. The functional core of the pilot factory is the Smart Production Data Platform. The platform operated by BENTELER IT fulfils three central tasks:

- The central provision of current and historical production data
- The execution of data analysis such as error detection
- The visualization and return of results to the user

For data provision, the machine controls were connected by means of standard interfaces (e.g. OPC-UA) and well over a thousand data sources have already been tapped in the plant. Signal changes in the range of less than 1 s are recorded, so that several million data points are recorded and made available every hour. In addition to real-time data, several years of historical data recordings can be accessed. These are necessary for the development and testing of data analysis methods, such as machine learning methods.

Dashboard usability and data interpretability are of prominent importance to ensure effective data visualization and decision support experience. Standard solutions like Grafana enable employees on the shopfloor to develop dashboards and individual displays independently. Individual machine data as well as the results of an anomaly detection are both available as data sources. The capability of the workforce to easily create alarms has been introduced in the decision workflow. The result is a significant time reduction in the response to unexpected events or even anticipation to failures. These new features also allow that out-of-range critical values or the frequent occurrence of anomalies can be reported immediately and addressed effectively to allow reduction of unplanned breakdowns. For further improvement of fault detection, employee feedback on the store floor by means of a decision support system is installed.

The production data platform, see Fig. 9, complies with the modular approach to Boost 4.0 big data pipeline development and open digital factory reference framework. It deploys modern technologies for container management, which allows the utilization of reusable software modules, for example for data provision, error detection, reporting or visualization. The individual modules can be flexibly combined to form new services, and a service can be transferred to another plant in just a few steps. The error detection for material handling systems developed in the Paderborn plant has already been tested in other BENTELER plants. The so-called micro-service architecture allows the fast and flexible development, adaptation and

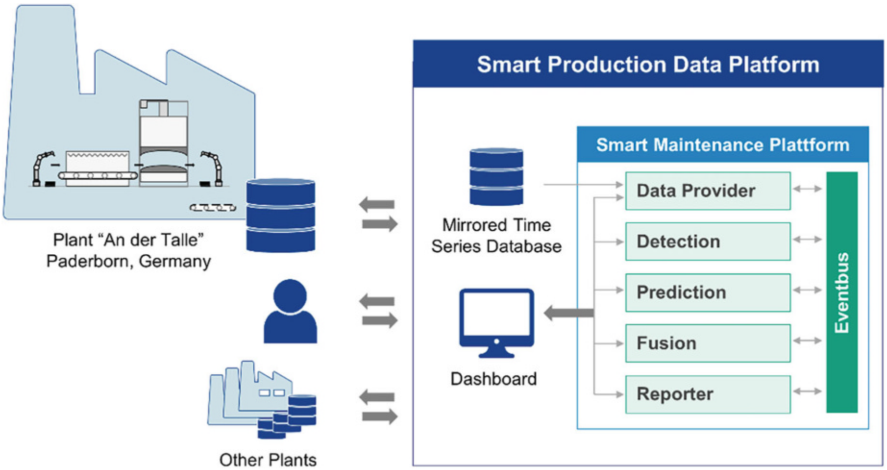

Fig. 9 Production data platform: A Smart Maintenance service accesses the provided data, uses modules for data analysis and visualizes the results in a dashboard (source: IEM/ATLANTIS Engineering)

testing of smart maintenance solutions. At the same time, it provides a future-proof architecture for other applications in production, such as process optimization or Smart Quality. The platform is already being used in other OWL research projects, for example ML4Pro [11]—Machine Learning for Production and its products.

3.3 Big Data-Driven Intelligent Maintenance Large-Scale Trial Performance Results

The implementation of the production data platform enables the deployment of software solutions that take advantage of Industry 4.0 technologies. One example is the Smart Maintenance Platform (SMP), which is able to monitor the machinery equipment of potentially all the BENTELER plants that are connected to the platform from a central remote location. Based on virtualization technologies, like Docker, and utilizing a micro-service architecture, SMP is able to *scale* its resources both vertically (e.g. adapt the system resources like CPU cores) and horizontally (e.g. deploy more instances in parallel of the Anomaly Detection micro-service), in order to cope with demanding data streaming scenarios.

Apart from the *scalability challenge* of the Big Data processing, SMP should also address the *transferability challenge*, in order to enable its application in different scenarios and use cases among the connected BENTELER plants. As already stated, the supervised learning-based monitoring approaches require the existence of fault data (i.e. machinery failures and errors), which in most of the crucial cases are rare due to preventive maintenance. However, SMP offers a set of Fault Detection tools, which utilize unsupervised learning approaches. Hence, only configuration over the data-intensive training of the supervised approaches is required in order to be applied. Of course, the great potential of the supervised predictive approaches is not neglected, as Fault Prediction tools are also offered by the platform, once enough fault data are collected by the Fault Detection tools and their training is feasible.

The performance of both the Fault Detection and Prediction approaches in terms of *Precision* (i.e. TP / (TP + FP)), *Recall* ((TP) / (TP + FN)) and *Accuracy* (i.e. (TP + TN) / (TP + TN + FP + FN)), where TP, TN, FP, FN are given by Table 1, is of special importance.

Table 2, depicts indicative results for both Fault Detection and Fault Prediction tools applied in the Paderborn plant analysing data of 1.5 years. The Fault Prediction

Table 1 The four outcomes of the data analysis tools

		Actual case	
		Fault	Normal
Predicted case	Normal	(False Negative) FN	(True Negative) TN
	Fault	(True Negative) TP	(False Negative) FP

Table 2 Fault Prediction and Detection results of the Hydraulic Press and the Material Handling use case in the Paderborn plant

		Precision	Recall	Accuracy
Hydraulic Press use case	Fault Prediction	0.5	1	0.93
	Fault Detection	0.99	1	0.99
Material Handling system	Fault Detection	0.93	0.95	0.99

was applied only in the Hydraulic Press use case as the behaviour of the Material Handling System was unpredictable. The results of the prediction approach should not be compared with the results from the detection approach as they are computed differently; however, the low precision of the prediction shows the difficulty of the approach to be trained properly as only three incidents occurred in 1.5 years.

Applying the Fault Prediction and Fault Detection tools to the Paderborn production line has already shown promising results that have the potential to remain at the same or even better levels, once the tools are adopted at a larger scale. The key performance indicators of interest for BENTELER from the business point of view are:

- Reduction in maintenance cost
- Reduction in MTTR (Mean Time To Repair)
- Increase in MTBF (Mean Time Between Failures)
- Increase in OEE (Overall Equipment Efficiency)

It should be mentioned that the application of the tools for certain equipment has already indicated the possibility of reducing the MTTR by 30% and of at least doubling the MTBF for certain types of failures.

3.4 *Observations and Lessons Learned*

The implementation of smart maintenance use cases posed not only technical challenges, but also challenges in project organization, e.g. communication with stakeholders within the company, knowledge management and its transfer between stakeholders and acceptance of developed solutions. In terms of domain and data understanding, using semi-formal models was a key to successful knowledge transfer. Constructing easy-to-understand, interdisciplinary models in joint workshops also increases acceptance and awareness for the involved stakeholders. The development of user-friendly and easily understandable dashboards allowed the demonstration of benefit of the smart production platform at shopfloor level. The utilization of reusable software modules facilitated the quick construction of a solution and transfer to other plants.

The implementation of a production data platform has proven to play a central role in the digitalization of BENTELER plants. It provides the basis for all data-driven use cases and data-driven decision-making: transparency about individual production machines as well as extensive production processes, monitoring and alerting, and advanced data analytics not only in smart maintenance, but also smart quality and process optimization. The decision to invest in the implementation of a production data platform thus is a complex matter, since it involves a comprehensive benefit analysis that is difficult to quantify. It is a mostly strategic decision, setting the roadmap for further approaches to factory operation and optimization.

4 Predictive Maintenance and Quality Control on Autonomous and Flexible Production Lines: The FCA Trial

4.1 Data-Driven Digital Process Challenges

The main challenges related to the pilot regard firstly the data management, starting from their collection, which can be difficult because of the different sensing systems implemented on the shopfloor production actors (e.g. accelerometers on AGVs and power meter on the laser cells). The presence of heterogeneous devices means the need to deal with specific communication protocols and different data acquisition speeds.

Another aspect, linked to the previous one, concerns the possible speed mismatch between production process, with the related data generation, and the information flow. As a consequence, one of the main challenges consists in the reduction of this time that has to be approximately equal to zero.

Then, an important challenge source is data protection, since security of the data is a crucial element of the pilot as we are dealing with the industrial field and especially with the production sector (e.g. production levels) and the quality sector (e.g. level of default). In particular, the management of data exposition to external providers on the cloud platform becomes very important, which necessitates careful data subdivision.

An additional field of action is represented by the communication between the industrial field and the cloud platform, because of the presence of security policies regarding the data flow which have to be respected, and that could represent a strong constraint for the pilot development.

Moving then to data utilization, the understanding, organization and use of the expansive datasets made available in new and better ways pursuing data uniformity and standardization across the entire product development lifecycle bring incredible challenges for data exploitation.

Then, regarding the data processing and analytics, we have that the entire process from the data acquisition, which sometimes could comprehend some edge pre-processing in order to reduce the volume of stored data, to data transfer and cloud analysis, in several cases has to be fast enough in order to enable near-real-time process feedback.

Finally, different challenges may come from the fields of data visualization and user interaction, since several end-users are considered in the pilot, from operators at the shopfloor level to maintenance operators at the information/operational level.

4.2 Big Data Manufacturing Process Value in Autonomous Assembly Lines in Automotive

The initial business scenario is about the implementation of the concept of autonomous production, where the traditional linear process is removed and mobile robots, such as Automated Guided Vehicles (AGVs); collaborative robots with vision capabilities; and fixed production cells collaborate together. In the traditional production processes, mobile robots have only 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 lead-time is crucial to ensure the stability and throughput of the production systems.

Planning, control, monitoring and maintenance of the mobile robots are required due to the fact that currently there is no specific approach to store and analyse data related to the missions of the vehicles, their wear-out and availability, taking into account the lead time for delivery and the uncertainty related to the interaction with the presence of human operators.

One of the main objectives is to ensure that the new technology is 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.

The autonomous assembly line aims to provide 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.) (Fig. 10).

AGVs are used to replenish and handle material or work-in-process between the different production islands, in particular the assembly and welding cells, and to/from the warehousing areas. Production actors are connected to the different production management platforms.

In the new scenario, production data coming from AGVs and laser cells are collected and enriched using FIWARE technology. Then they are sent through MindConnect technology and stored in a data lake on cloud provided by SIEMENS MindSphere Platform.

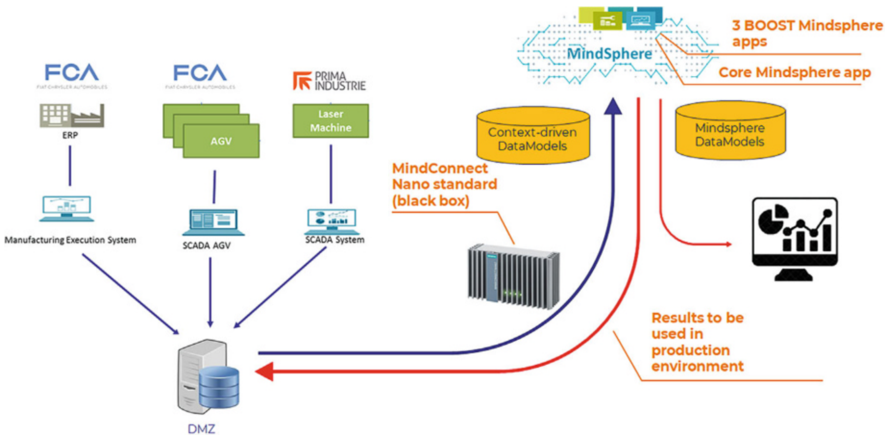

Fig. 10 FCA trial big data pipeline for autonomous assembly lines (AGV and Laser cells)

Different algorithms elaborate the production data in order to monitor the quality of the produced components, detect malfunctions enabling the definition of a maintenance schedule and optimizing the allocation of production missions. Besides, the different data are made accessible to external service providers, in order to enable the development of innovative applications based on proprietary data. To this end, and to ensure data privacy and security, open data models have been developed and IDS technology (e.g. IDS connector) has been implemented.

4.3 *Implementation of Big Data Solutions*

The pilot development began with a prototype application, which gathered sensors data from an assembly cell, replenished by an AGV, located in the Campus Melfi shopfloor. Data were collected from the machines to a central database, and they were visualized by the prototype application through a dashboard.

Successively, it has evolved into an industrial experimentation site, which started from the results provided by the prototype application in order to progressively develop and test the complete pilot architecture. It began with a first phase, in which data were gathered from different sources into MindSphere [12], which implemented the IDS architecture [13] and has been structured on the basis of the data sovereignty principles. The data sources were represented by an AGV owned by FCA and located in the Melfi Campus, and a laser machine owned by PRIMA and located within the Prima's labs. MindSphere hosted on the cloud and data from the shopfloor were exposed to external service providers. Specific APIs, called Mindlib [14], were used to send data from the source systems (data provider) to MindSphere Services Platform (data consumer). A datamodel has been created and used to set

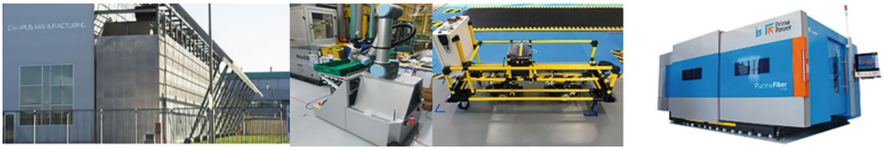

Fig. 11 FCA Boost 4.0 Melfi Campus Experimental Site and several pilot production actors

up the platform, to be able to collect the data. A visualization App (MindApp) displayed the data (Fig. 11).

Then, in the second phase a scenario of interaction between the robots (AGVs) and the production cells (fixed machines from PRIMA) was implemented and tested. Within this scenario, the manufacturing capability w granted by the correct functioning of both the robots and the fixed machines. Three specific apps were defined in the MindSphere environment and mostly the first two, the PRIMA Fleet Management App which monitors the main parameters trends analysing and correlating different types of data and the Smart Scheduling App which optimizes the mission allocation to the different production actors, were developed. The number of data sources from AGVs and Laser machines were widened and a DMZ (demilitarized zone) was set up in CRF in order to permit the interface between the industrial environment and MindSphere.

Lastly, during the third phase, the pilot was extended to the full industrial scale and so the architecture was adapted to the final number of data sources and data amount, the connector to the Fiware Orion Context Broker was developed and inserted in the data flow system and an app for anomaly detection using data coming the AGVs was finalized.

4.4 Large-Scale Trial Performance Results

The implementation of an Industrial IoT Data Space based on the MindSphere platform has allowed FCA to develop a number of MindApps with the collaboration of external service providers. This approach has delivered three big-data driven innovative services that are currently being operated at the Melfi Campus:

- MindApp for production optimization
- MindApp for welding quality control
- MindApp for AGV anomaly detection (Fig. 12)

These three apps are significantly improving the performance and flexibility of the autonomous assembly lines, adapting the production scheduling to real-time sensitivity to production quality variation and asset maintenance needs to allow zero unplanned breakdown.

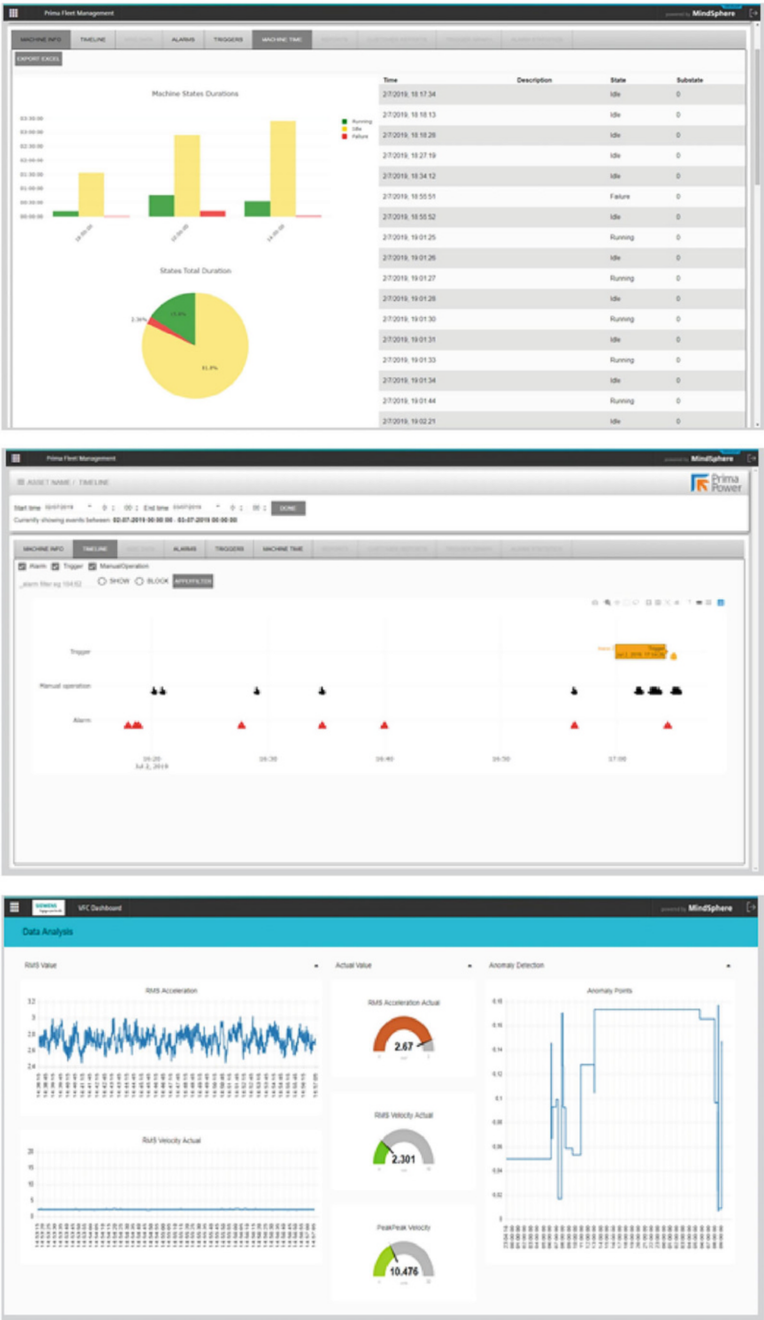

Fig. 12 MindApps dashboards developed by PRIMA and third-party developers at FCA Melfi Campus Experimental Site based on the IIoT MindSphere industrial data space

4.5 Observations and Lessons Learned

The pilot development and implementation presented some barriers that had to be overcome. Regarding the software development, the main effort has been represented by the connectivity aspects. In order to make the applications work properly, it was necessary to have data in the correct aggregation and format as required by MindSphere. Therefore, the development of connectors and format converters was the most expensive part, along with the coding and application deployment, in terms of resources and effort. Moving then to the data flow architecture, the main issue here was the choice and the development of a solution which allowed the exposition of the industrial data to external partners avoiding to put in danger the security of the entire company's internal network. The identification and the development of the solution, which consists in a Demilitarized Zone, required a huge effort in collaboration with the IT and security departments in order to, on one side, respect all the company policies and, on the other, meet all the project requirements that would have led then to the development of the data transfer infrastructure.

5 Conclusions

This chapter has presented the advanced manufacturing processes and big data-driven algorithms and platforms leveraged by the Boost 4.0 big data lighthouse project that allow improved digital operations within increasingly automated and intelligent shopfloors. It has demonstrated how three different companies have been able to implement three distinct, open, yet sovereign cross-factory data spaces under a unified framework, i.e. Boost 4.0 big data reference architecture and Digital Factory Alliance (DFA) [15] service development framework. Philips has provided evidence of the significant benefits that data spaces and integrated data pipelines can bring to their Drachten brownfield production lines in terms of implemented increasingly predictive quality control and fact-based automated decision support processes. BENTELER Automotive, has equally demonstrated the benefits of a modular and data-space approaches to deliver high cross-factory transferability of smart maintenance 4.0 services from their factory in Padeborn, all based on the use of advanced software containerization and virtualization as well as open source technology for the implementation of data spaces and data pipelines. Finally, FCA has demonstrated the benefits and challenges that the operation of Industrial IoT data spaces supported by MindSphere entail to support the implementation of flexible, modular autonomous assembly cells. FCA has demonstrated how the implementation of such data spaces in their Melfi Campus facilities based on open APIs allows not only a better integration of the shopfloor assets but also opens up the opportunity for the development of high-value customized services and data-driven apps that positively impact the performance of the digital shopfloor and allow a more resilient and adaptive scheduling of production. The chapter has shown that maintenance performance improvements (main time between failures) can be improved by 600%, overall equipment efficiency (OEE) by 14% and production efficiency by 10%. These figures are close to those estimated by literature studies and can be achieved by means of adopting a unified big data approach provided by the Boost 4.0 reference model, the implementation of industrial data spaces and the realization of advanced decision support dashboards that reduce the time to decision and action data. Boost 4.0 has demonstrated that industry can cost-effectively implement effective means for data integration, even in brownfield production lines with significant legacy equipment.

Acknowledgements This research work has been performed in the framework of the BOOST 4.0 Big Data Lighthouse initiative, a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 780732. This data-driven digital transformation research is also endorsed by the Digital Factory Alliance (DFA) www.digitalfactoryalliance.eu

References

1. European Economic and Social Committee and the Committee of the Regions. (2020). *A European strategy for data*. European Commission. Available at: https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf. Accessed January 2021
2. World Economic Forum. (2016). Digital transformation of industries demystifying digital and securing $100 trillion for society and industry by 2025.
3. Boost 4.0. https://www.boost40.eu. Accessed January 2021.
4. Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.). (2017). *European big data value strategic research & innovation agenda*. Big Data Value Association.
5. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O'Sullivan, B., & Smeulders, A. (Eds.). (2020, September). *Strategic research, innovation and deployment agenda—AI, data and robotics partnership*. Third release. Brussels. BDVA, euRobotics, ELLIS, EurAI and CLAIRE.
6. D2.5 Boost 4.0 reference architecture specification. Available at: https://boost40.eu/wp-content/uploads/2020/11/D2.5.pdf. Accessed January 2021.
7. Kepware. https://www.kepware.com. Accessed January 2021.
8. Inmation. https://www.inmation.com. Accessed January 2021.
9. Roland Berger GmbH. (2017). Predictive maintenance. Service der Zukunft—und wo er wirklich steht. Studie im Auftrag des VDMA.
10. Henke, M., Heller, T., & Stich V. (Eds.). (2019). *Smart maintenance—Der Weg vom Satus quo zur Zielvision (acatech STUDIE)*. utzverlag GmbH.
11. ML4Pro. https://www.iem.fraunhofer.de/de/referenzen/forschungsprojekte/benteler-ml4pro2.html. Accessed January 2021.
12. Mindshpere. https://siemens.mindsphere.io. Accessed January 2021.
13. International Data Spaces Association. (2019). Reference architecture model. Online. Available at: https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf. Accessed January 2021.
14. MindLib. https://community.sw.siemens.com/s/article/data-structure-and-data-upload-to-mindsphere-via-mindlib. Accessed January 2021.
15. Digital Factory Alliance (DFA). https://digitalfactoryalliance.eu. Accessed January 2021.