

# Data-Driven Artificial Intelligence and Predictive Analytics for the Maintenance of Industrial Machinery with Hybrid and Cognitive Digital Twins



Perin Unal, Özlem Albayrak, Moez Jomâa, and Arne J. Berre

**Abstract** This chapter presents a Digital Twin Pipeline Framework of the COGNITWIN project that supports Hybrid and Cognitive Digital Twins, through four Big Data and AI pipeline steps adapted for Digital Twins. The pipeline steps are Data Acquisition, Data Representation, AI/Machine learning, and Visualisation and Control. Big Data and AI Technology selections of the Digital Twin system are related to the different technology areas in the BDV Reference Model. A Hybrid Digital Twin is defined as a combination of a data-driven Digital Twin with First-order Physical models. The chapter illustrates the use of a Hybrid Digital Twin approach by describing an application example of Spiral Welded Steel Industrial Machinery maintenance, with a focus on the Digital Twin support for Predictive Maintenance. A further extension is in progress to support Cognitive Digital Twins includes support for learning, understanding, and planning, including the use of domain and human knowledge. By using digital, hybrid, and cognitive twins, the project's presented pilot aims to reduce energy consumption and average duration of machine downtimes. Data-driven artificial intelligence methods and predictive analytics models that are deployed in the Digital Twin pipeline have been detailed with a focus on decreasing the machinery's unplanned downtime. We conclude that the presented pipeline can be used for similar cases in the process industry.

**Keywords** Digital Twin · Data-driven Artificial Intelligence · Big Data pipeline · AI pipeline · Hybrid Digital Twin · Cognitive Digital Twin · Predictive Analytics · Predictive Maintenance

---

P. Unal (✉) · Ö. Albayrak  
Teknopar Industrial Automation, Ankara, Turkey  
e-mail: [punal@teknopar.com.tr](mailto:punal@teknopar.com.tr)

M. Jomâa  
SINTEF Industry, Oslo, Norway

A. J. Berre  
SINTEF Digital, Oslo, Norway

# 1 Introduction

The COGNITWIN project,<sup>1</sup> “Cognitive Plants Through Proactive Self-Learning Hybrid Digital Twins”, is a 3-year project in the Horizon 2020 program for the Process Industry, from 2019 to 2022. The project focuses on using Big Data and AI for the European Process Industry through the development of a framework with a toolbox for Hybrid and Cognitive Digital Twins.

This chapter describes the Digital Twin approach and demonstrates its application to one of the project’s six use case pilots: Spiral Welded Machinery (SWP) in the steel pipe industry. The chapter relates to the data spaces, platforms, and “Knowledge and Learning” cross-sectorial technology enablers of the AI, Data and Robotics Strategic Research, Innovation & Deployment Agenda [32]. The chapter is organised as follows:

Section 2 presents the *Digital Twin Pipeline Framework* from the COGNITWIN project supporting Digital, Hybrid and Cognitive Digital Twins, through the four pipeline steps. Section 3 briefly introduces the state of the art and state of the practice in maintenance of industrial machinery. Section 4 presents the pilot Application of *Maintenance of Industrial Machinery for Spiral Welded Steel*, focusing on the Digital Twin support for Predictive Maintenance for Machines in the Welded steel plant/factory. Section 5 describes the *Big Data and AI Technology* selections of the Digital Twin system applied for the steel industry case related to the different technology areas in the BDV Reference Model. Section 6 details COGNITWIN Digital Twin Pipeline Architecture and the Platform Developed and presents the four pipeline steps of the COGNITWIN Digital Twin Pipeline realised in the context of the Spiral Welded Steel pilot case. Section 6.1 explains how *Digital Twin Data Acquisition and Collection* is taking place from factory machinery and assets with connected devices, and controllers through protocols and interfaces like OPC UA and MQTT. Section 6.2 exemplifies *Digital Twin Data Representation* in various forms, based on the sensor and data sources connections involving event processing with Kafka and storage in relevant SQL and NoSQL databases combined with Digital Twin API access opportunities being experimented with, such as the Asset Administration Shell (AAS). Section 6.3 presents *Digital Twin Hybrid (Cognitive) Analytics* with AI/Machine learning models based on applying and evaluating different AI/machine learning algorithms. This is further extended with first-principles physical models—to form a Hybrid Digital Twin with examples of data and electrical and mechanical models for a DC motor to support predictive maintenance. Section 6.4 describes the pipeline step for Digital Twin Visualisation and Control, including the use of 3D models and dashboards suitable for interacting with Digital Twin data and further data access and system control through control feedback to the plant/factory. Finally, the conclusion in Sect. 7 presents a summary

---

<sup>1</sup> <http://cognitwin.eu/>

of this chapter’s contributions and the plans for future improvement of technologies and use case pilots in the COGNITWIN project.

## 2 Digital Twin Pipeline and COGNITWIN Toolbox

Many of the advancements and requirements related to Industry 4.0 are being fulfilled by the use of Digital Twins (DT). We have in earlier papers introduced our definitions for Digital, Hybrid, and Cognitive Twins [1–3]—which also aligns with definitions of others [4, 5]: “A DT is a digital replica of a physical system that captures the attributes and behaviour of that system” [6]. The purpose of a DT is to enable measurements, simulations, and experimentations with the digital replica to gain an understanding of its physical counterpart. A DT is typically materialised as a set of multiple isolated models that are either empirical or first-principles based. Recent developments in artificial intelligence (AI) and DT bring more abilities to the DT applications for smart manufacturing.

A hybrid twin (HT) is a set of interconnected DTs, and being an extension of a HT, a cognitive twin (CT) is a self-learning and proactive system [6]. The concepts of HT and CT are introduced as elements of the next level of process control and automation in the process and manufacturing industry. In the COGNITWIN project, we define an HT as a DT that integrates data from various sources (e.g., sensors, databases, simulations, etc.) with the DT models, and applies AI analytics techniques to achieve higher predictive capabilities, while optimising, monitoring, and controlling the behaviour of the physical system. A Cognitive Twin (CT) is defined as an extension of HT incorporating cognitive features to enable sensing complex and unpredicted behaviour and reason about dynamic strategies for process optimisation. A CT will combine expert knowledge with the power of HT.

We have adopted the Big Data and AI Pipeline that have been described in [7] and specialised this for the context of Digital Twins as shown in Fig. 1.

The proposed pipeline architecture starts with data acquisition and collection to be used by the DT. This step includes acquiring and collecting data from various sources, including streaming data from the sensors and data at rest.

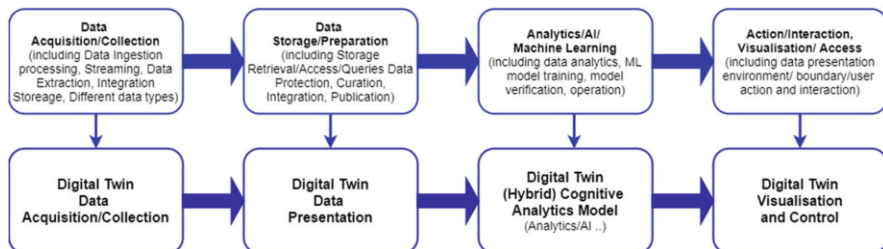


Fig. 1 Big Data and AI Pipeline architecture—applied for Digital Twins

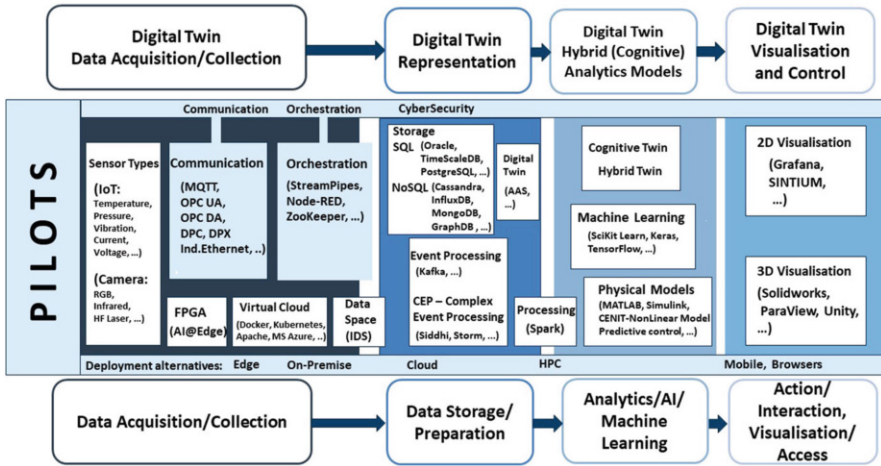


Fig. 2 Cognitive Twin Toolbox with identified components for the various pipeline steps

Following the data acquisition and collection, the next step is the DT data representation in which the acquired data is stored and pre-processed. The DT (Hybrid) Cognitive Analytics Models step of the pipeline enables integration of multiple models and the addition of cognitive elements to the DT through data-analytics. Finally, the DT Visualisation and Control step of the pipeline provides a visual interface for the DT, and it provides interaction between the twin and the physical system.

Figure 2 shows various components in the COGNITWIN Toolbox that can be selected in order to create Digital Twin pipelines in different application settings.

The COGNITWIN Toolbox is being used to create operational Digital Twin pipelines in a set of use cases as follows:

- Operational optimisation of gas treatment centre (GTC) in aluminium production, with support for the recommendation of optimal operating parameters for adsorption based on real-time data gathered about conditions such as the pressure, temperature, humidity, etc., from sensors.
- Minimise health and safety risks and maximise the metallic yield in Silicon (Si) production to provide best estimates of when the furnace can be emptied to the ladle for further operations.
- Real-time monitoring of finished steel products for operational efficiency with an ability to react on its own to situations requiring intervention, thus stabilising the production process further.
- Improving heat exchanger efficiency by predicting the deposition of unburnt fuel mixtures, ash, and other particles on the heat-exchanger tubes based on both historical practices and real-time process.

In the following, we illustrate the approach for creating an operational Digital Twin pipeline for the use case on predictive maintenance for Steel Pipe Welding. We plan to apply it to improve the Steel Pipe Welding process further as follows:

- Life cycle optimisation of Spiral Welded Machine (SWP) in steel pipe production, where CT of the SWP monitors the condition and health of the machinery, offers early warnings, and suggests optimised predictive maintenance plans for the machinery based on real-time data gathered from sensors such as the pressure, temperature, vibration, etc., and alarm and status information.
- Improving operational performance of the production process by predicting and identifying the optimal operating parameters based on both historical practices and real-time process and thus improving the overall productivity of the plant.
- Improving energy consumption efficiency by monitoring and predicting the energy analyser and operational parameters based on both historical practices and real-time process.
- Enhanced utilisation of computing infrastructure with virtual machines and containerisation technologies to achieve optimised RAM and CPU usage.
- Minimise health and safety risks and maximise the human operator performance by early warning of machine and system problems.
- Real-time monitoring of parameters like pipe diameter, pitch angle, belt width, production speed, pipe diameter and wall thickness for semi-finished and finished steel products for ensuring operational efficiency and stabilising the production process.

### 3 Maintenance of Industrial Machinery and Related Work

Maintenance in the production industry has always been an important building block providing essential requirements, such as cost minimisation, prolonged machine life, and increased safety and reliability. On the other hand, Predictive Maintenance (PM) has been a popular topic of research for decades with hundreds of papers having been published in the area. Since machine learning techniques came into prominence in the field with the emergence of industry 4.0, PM has become an even more important area of interest [8].

There are three maintenance strategies. The first is Reactive Maintenance (RM) in which little or no maintenance is undertaken until the machine is broken. The second type is Preventive Maintenance, which is based on the repair or replacement of equipment on a fixed calendar schedule regardless of their condition. This approach has benefits over RM, but it can lead to the unnecessary replacement of components that may still be in good working condition, resulting in increased downtime and waste. The last and most recent method is Predictive Maintenance, in which the main goal is to precisely estimate the Remaining Useful Life (RUL) of machinery based on various readings of sensor data (heat, vibration, etc.).

Unplanned downtime of machinery often results in economic losses for companies. Thus, predicting timely machine needs for maintenance can result in financial gain [9] and reduce unnecessary maintenance operation due to the preventive approach [10]. Another advantage of PM is cost minimisation, including minimising fatal breakdowns and reducing certain components' replacement, which is closely related to the other benefits [11].

There are many uses for PM in industrial application. A large amount of event data related to errors and faults in the internet of things (IoT) and digital platforms are continuously collected. An event records the behaviour of an asset, and by nature it comes in the form of data-streams. Event Processing is a method for processing streams of events to extract useful information, and complex event processing (CEP) is used for detecting anomalies and for predictive analytics [12]. Although event processing has been a paradigm for data processing for nearly three decades, there have been recent advancements in the last decade due to novel applications by the IoT and machine learning technologies. The presence of large amounts of streaming sensor data that can be widely generated is another reason for the growing interest in event processing. Predictive analytics uses the streams of data to make predictions of future events.

Making use of Predictive Analytics and CEP together provides synergy in PM performance [13]. Lately, there has been increased usage of event processing platforms that use open source technologies for Big Data and stream processing. Sahal, Breslin and Ali used an open-source computation pipeline and showed that the aggregated event-driven data, such as errors and warnings, are associated with machine downtime and can be qualified as predictive markers [14]. Calabrese et. al. performed equipment failure prediction from event log data based on the SOPHIA architecture, which is an event-based IoT and machine learning architecture for PM in Industry 4.0 [15]. Aivaliotis, Georgoulas and Chryssolouris investigated PM for manufacturing resources by utilising physics-based simulation models and the DT concept [16].

This research aimed to approach PM based on an open-source event processing platform and allow for the accurate prediction of time to failure to increase machine availability. Data-driven models are accompanied by a machine learning library and the DT concept to analyse the components of a machine's health status. The DT enables the platform under study to be a PM system, rather than a predictive analytics system. The DT concept used together with the open-source event-driven platform is detailed in [6], and the DT developed together with the platform is presented in [2].

Maintenance approaches that can monitor equipment conditions for diagnostic and prognostic purposes can be grouped into three main categories: statistical, artificial intelligence, and model based [17]. The model-based approach requires mechanical and theoretical knowledge of the equipment to be monitored. The statistical approach requires a mathematical background, whereas, in artificial intelligence, data are sufficient; thus, despite the challenges in the data science pipeline (data understanding, integration, cleaning, etc.), the last approach has been increasingly applied in PM applications.

In this study [18], the authors defined three types of PM approaches: (1) The data-driven approach, also known as the machine learning approach, uses data; (2) the model-based approach relies on the analytical model of the system; and (3) the hybrid approach, combining both methods. Dalzochio et. al. evaluated the application of machine learning techniques and ontologies in the context of PM and reported the application areas as fault diagnosis, fault prediction, anomaly detection, time to failure, and remaining useful life estimation, which refer to the several stages of PM [19].

According to [20], artificial neural networks (ANNs) are widely used for PM purposes. Carvalho et. al. noted that random forest (RF) was the most used method in PM applications [21]. Compared to the other machine learning methods in PM applications, RF was found to be more complex and it took more computational time [21]. In their review, the authors also noted that ANNs were one of the most common and applied machine learning algorithms in industrial applications since they were only based on previous data, requiring minimum expert knowledge, and the need for coping with challenges of data science pipeline.

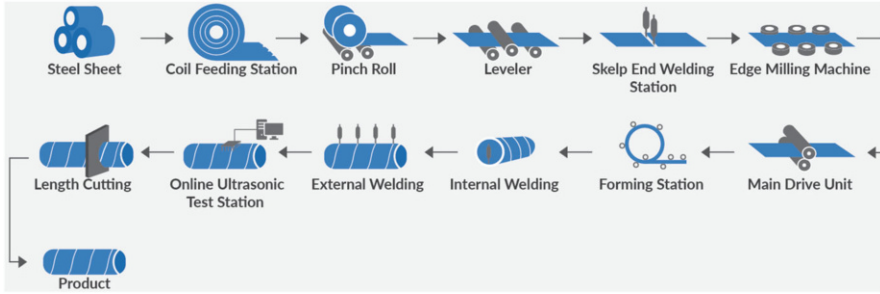
Another study [22] evaluated the performance of k-Nearest neighbour (kNN), back-propagation feed-forward neural network (FFNN), DecisionTree, RF, support vector machine (SVM), and naïve Bayesian and assessed the results for time-series prediction. Rivas et. al. used the long short-term memory (LSTM) recurrent neural network (RNN) model, another type of ANN capable of using memory, for failure prediction [23]. Kanawaday and Sane discussed the use of autoregressive integrated moving average to predict possible failures and quality defects [24]. They basically used this method to predict future failure points in the data series for the diagnosis of machine downtimes.

Another research [15] presented an architecture that used tree-based algorithms to predict the probability of a failure where the gradient boosting machine generated the models that obtained the best results for classification when compared to the distributed RF models and extreme gradient boosting models. The use of deep learning algorithms is another promising area in PM. In this context, [25] used convolutional neural networks in performing the task of extracting features, and [26] used deep neural networks to reduce the dimensionality of data.

## 4 Maintenance of Spiral Welded Pipe Machinery

Figure 3 presents the Spiral Welded Steel (SWP) pipe production process. The SWP pilot of the COGNITWIN project is one of the six pilots that aims at enabling predictive maintenance (PM) for the SWP machinery presented in Fig. 4. NOKSEL is one of the use case partners of the COGNITWIN project [27]. The figure shows the machinery used in NOKSEL's facilities, and the process followed by this machinery to produce steel pipes. With the Digital Twin-supported condition monitoring platform, an infrastructure that aims to analyse the operational and





**Fig. 3** Spiral welded steel pipe process



**Fig. 4** SWP Machinery and process for which PM is developed

automation data received from sensors and PLC/SCADA will be used for PM, which will help increase the overall equipment performance.

In the steel pipe sector, operations run on a 24/7 basis. Due to the multi-step and interdependent nature of the production process, a single malfunction in one of the work stations can bring the whole production process to a halt. Thus, the cost of machine breakdown is very high. Under the COGNITWIN project scope, the Digital Twin is developed for the production process of Spiral Welded Steel Pipe machinery (SWP). The goal is to make use of developed models and analyse multiple sensors' data streams in real-time and enable predictive maintenance to reduce downtimes by Digital Twins. The main targets to be achieved are:

- 10% reduction in energy consumption
- 10% reduction in the shifted average duration of downtimes



A DT on NOKSEL's production process of Spiral Welded Steel Pipes (SWP) collects, integrates, and analyses multiple sensors' data. CT will be built upon DT with an aim to autonomously detect changes in the process and to know how to respond in real time to the constantly changing scenario with minimal human intervention. The first iteration of DT and CT systems will be an HT where human input will be required to take action based on the CT system's feedback. The CT will have cognitive capabilities by using real-time operational data to enable understanding, self-learning, reasoning, and decisions.

All the parameters in the first-order model are set as same as the real ones and expert knowledge is integrated into the model for HT. The data collected from the plant and the results of the simulation model are compared to ensure consistency in an iterative process. The model is modified, and simulation is repeated iteratively until the difference between the simulation results and the data retrieved is negligible. The collected data are used to train machine learning models and make predictions. In the next step, the predictions from the data-driven model will be taken as the observation value of the hybrid algorithms to adjust the theoretical values. The theoretical values coming from first-order models and data coming from the digital platform will be fused by hybrid algorithms like Kalman filtering and particle filtering algorithms.

In the NOKSEL pilot case, CT will introduce improved decision-making by integrating human knowledge into decision-making process. The anomalies, alarms, and early warnings of machine and system problems will be tackled by CT, and the decision-making process will emulate the experienced human operator with an embedded knowledge base. CT will augment expert knowledge for unforeseeable cases on HT and DT. The human operator's experience is added to process knowledge model and physics-based models with parametric values as well as thresholds and causality relations. Expert knowledge on the causes of breakdowns is collected with the series of problematic operations and the initial causes that trigger the successive reactions.

Data Analytics is deployed to extract knowledge from data. Machine learning techniques are used for this purpose, and a machine learning library is built. In the top layer, Data Visualisation and User Interaction, an advanced visualisation approach is provided for improved user experience with low latency. In vertical layers Cybersecurity and Trust is supported by IDS Security component and the Communication and Connectivity layer is composed of REST API, Kafka [28], MQTT [29], and OPC [30] components.

In Fig. 5, relevant components from the COGNITWIN toolbox have been used in the creation of a DT pipeline for the Steel Pipe Welding. The platform is aimed to be used for use cases where continuous availability, high performance, flexibility, robustness, and scalability are critical.



## 5 Components and Digital Twin Pipeline for Steel Pipe Welding

The Digital Twin pipeline components are in the following mapped to the BDV Reference Model of the Big Data Value Association [31], which serves as a common reference framework to locate Big Data technologies on the overall IT stack. It has been presented in Fig. 6 and detailed in BDVA SRIA [32].

Things, assets, sensors, and actuator (IoT, CPS, edge computing) layer contain PLC as the main source of data, as well as the control units containing sensor data, alarms, and states of assets. In the Cloud and High-Performance Computing layer, the Big Data processing platform and data management operation are supported by the effective use of a private cloud system and computing infrastructure. Docker [33] is used here as a packaging and deployment methodology to easily manage the variety of the underlying hardware resources efficiently.

To meet system needs, Data Management is handled to collect and store raw data and manage the transformation of these data into the required form. The protection of these data is handled via privacy and anonymisation mechanisms like encryption, tokenisation and access control in the Data Protection layer. In the Data Processing Architecture layer, an optimised and scalable architecture is developed for the analytics of both batch and stream processing via SIMATIC and Spark [34] stream processing. Data clean-up and pre-processing are also handled in this layer.

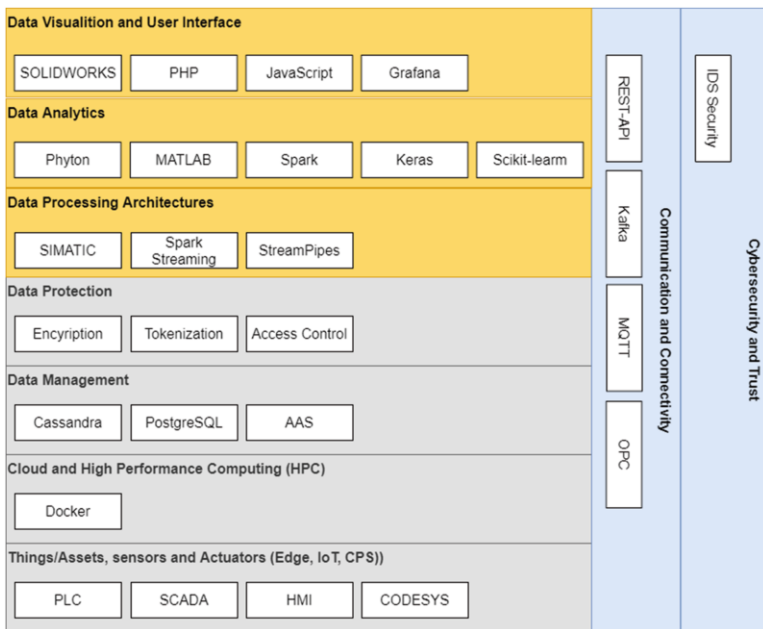


Fig. 6 The architecture aligned with BDVA Big Data Value Reference Model

## 6 COGNITWIN Digital Twin Pipeline Architecture

The four-step Digital Twin pipeline also has a mapping to the technical areas in the SRIDA for AI, Data and Robotics Partnership [45] as follows: Digital Twin Data Acquisition and Collection relate to enablers from *Sensing and Perception* technologies. Digital Twin Data Representation relates to enablers from *Knowledge and Learning* technologies, and also to enablers for *Data for AI*. Hybrid and Cognitive Digital Twins relate to enablers from *Reasoning and Decision*. Digital Twin Visualisation and Control relate to enablers from *Action and Interaction*. These four steps are described in more detail in the following.

### 6.1 Digital Twin Data Acquisition and Collection

IIoT refers to IoT technologies used in industry, and it has been the primary building block of the systems facilitating the convergence and integration of operational technology (OT) and information technology (IT) for gathering data from sites [35]. This section details the IIoT system used for data acquisition and collection.

Figure 7 shows the hardware topology of the previously existing system. With the Digital Twin-supported condition monitoring platform to be developed, and infrastructure that aims to analyse the operational and automation data received from sensors and PLC/SCADA will be used for PM, which will help increase the overall equipment performance.

The topology of the infrastructure established is shown in Fig. 7. Communication between these two topologies is provided with the industrial communication protocol PROFINET, and the two structures will communicate with each other. Data required from the existing structure can be obtained using the existing controller. Figure 8 presents the added hardware topology.

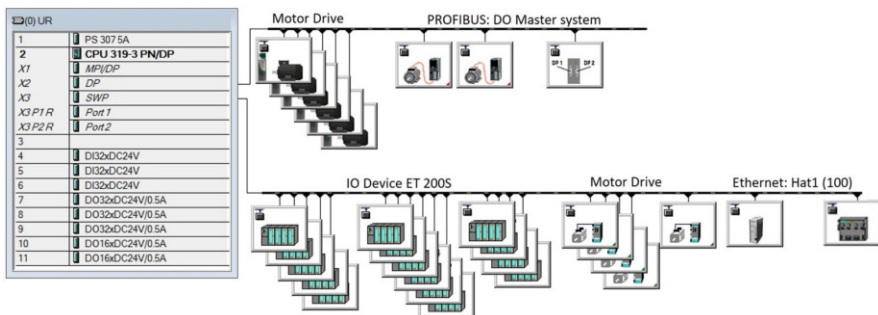


Fig. 7 Existing hardware topology

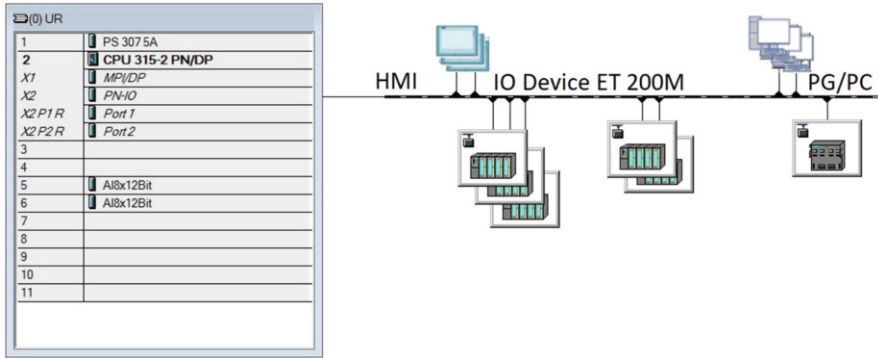


Fig. 8 Added hardware topology

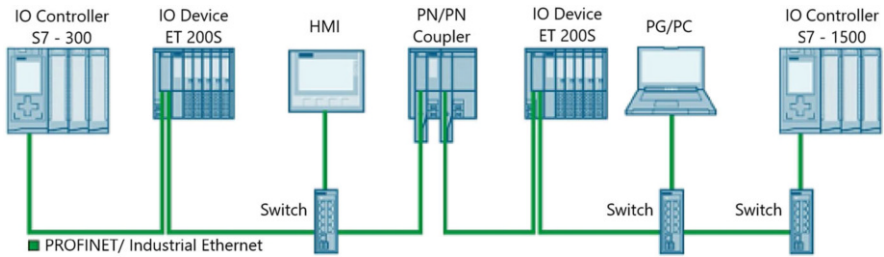


Fig. 9 Coupling of the PROFINET subnets with the PN/PN Coupler

The current PLC model used for process control is S7 300. The operation details of the components; status information; process information, such as speed and power; production details; and system alarms are kept on this PLC while the newly added sensor data and alarms will be located in the S7 1500 PLC. The existing PLC data will be transferred to the S7 1500 PLC through the PN/PN Coupler module, allowing all data tracking to be carried out over the new PLC.

The PN/PN Coupler module provides the simple connection of two separate PROFINET networks. The PN/PN Coupler enables data transmission between two PROFINET controllers. The output data from one network becomes the input data of the other. For data transfer, additional function blocks are not required and the transfer is realised without latency. For the new sensors added to the system not to affect the existing process, a new PLC is employed and the controls are implemented over it. The communication structure between the PLCs is designed using the PN/PN Coupler module as shown in Fig. 9.

PLC transmits the data it receives from the sensors to OPC, which then transfers the data to the platform via MQTT. The received data is transmitted to Kafka, which passes it on to the Cassandra [36] database to be stored for further processing or later access.

## 6.2 *Digital Twin Data Representation*

DT representation step follows the data acquisition/collection step. In this step, data collected in the data acquisition step is stored in the information domain, and the stored data is used by the business domain.

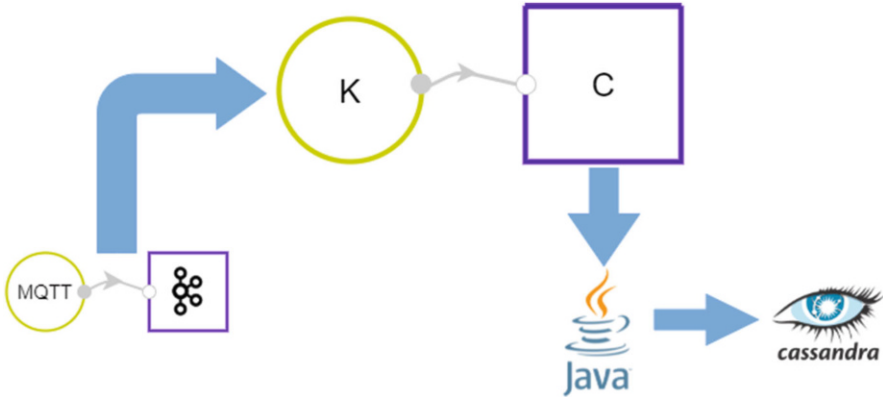
The data obtained from the sensors, such as temperature, pressure and vibration, voltage, and current, are transmitted to MQTT over OPC in the first tier, and then to Kafka in the JSON format. Apache Kafka is a data streaming platform developed specifically to transmit real-time data with a low error margin and short latency. Kafka achieves superior success in systems with multiple data sources, such as sensor data, and reduces the inter-system load. It has an integration that can also process Big Data coming from sensors operating at high frequencies.

Instant data received by Kafka is transmitted to the Python [37]-based server, where the attribute extraction process begins. Incremental principal component analysis (PCA), which is the most well-known method used in the Big Data flow, applies PCA stages to the instantaneous data using data in a certain window range, and thus large data that cannot fit into the memory can also be processed effectively. PCA performs dimensional reduction by making the incoming high-dimensional data low-dimensional, providing more accurate results for machine learning, and therefore it is frequently used for categorisation problems.

A fully asynchronous communication structure with the event-bus method is used for the transmission of data collected from the source with OPC. Data transmission is provided in the JSON format. In the architecture managed based on Microservices, Cassandra is used as the NoSQL database with a database presented as a log file to users. Cassandra is a database that provides continuous availability, high performance, and scalability. PostgreSQL [38], a relational database (RDBMS), is used by the interface program that provides user interaction to display time-series data in real time.

A total of 120 sensor values are monitored on the SWP machine to capture data on temperature, vibration, pressure, current, oil temperature, and viscosity. A value is taken every 10 ms from the vibration sensors, once every 100 ms from the current sensors, and every 1000 ms from the temperature and pressure sensors plus alarm and status fields. The SWP machinery has a total of 120 sensor values, 122 alarms, and 175 status data which create 11 GB of incoming data in 1 day (24 h).

To organise the collected data in a Digital Twin structure, we have analysed several emerging Digital Twin open source-initiatives and standards, as reported in the COGNITWIN project survey paper “Digital Twin and Internet of Things-Current Standards Landscape” [3]. Based on this, we have selected to use the Asset Administration Shell (AAS) for further Digital Twin API development. AAS was developed by Platform Industry 4.0, and similar to DT, AAS is a digital representation of a resource [3]. Descriptions of AAS can be serialised using JSON, XML, RDF, AutoML and OPC UA [39]. To realise COGNITWIN vision, the AAS model and APIs are not sufficient. For COGNITWIN, the models and the components using different technologies should be reusable. For this purpose,



**Fig. 10** An example pipeline created in Apache StreamPipes

it is decided to utilise Apache StreamPipes [40] for the IIoT. StreamPipes is expected to provide an environment to host different models, components, and services, and to orchestrate them. Figure 10 presents a sample pipeline generated in Apache StreamPipes. As shown, MQTT data is retrieved by Kafka and stored in the Cassandra database in a pipeline. The main purpose of the developed pipeline is to create a path from sensory data to the neural network output, depicting the state of the tool, through several pipeline elements. Besides, this pipeline triggers notification when the value of a particular property goes above a certain threshold and shows results. The below sample shows the pipeline elements regarding data collection on data storage.

### 6.3 Hybrid and Cognitive Digital Twins

In the Machine Learning Library (MLL) module, different machine learning algorithms are applied through the incremental PCA stage to detect anomalies. Prediction results are produced using various machine learning libraries. First, is Spark MLlib produced entirely by Spark, which uses Spark's engine optimised for large-scale data processing. Keras library utilises TensorFlow, and is used for deep learning. The LSTM algorithm of this library is utilised. This open-source neural network library makes it simpler to work with artificial neural networks through its user interface facilities and modular structure. The Scikit-Learn [41] library is another open-source machine learning library that contains several algorithms for regression, classification, clustering. We used algorithms like RF, GBT, LSTM, SVM, KNN and multi-layer perceptron (MLP) from Scikit-Learn library for data modeling and prediction.

The MLL module is used for comparing the different machine learning models. When setting up a machine learning model, it is difficult to predict which model



architecture will provide the best results. The parameters that can affect the model architecture are called hyper-parameters. For each machine learning algorithm used, hyper-parameters tuning is performed by comparing the previously determined success criteria and selecting the best result combination by looking at the results obtained by testing possible combinations of the values of the hyper-parameters in a certain range. For each ML algorithms used, in addition to different parameters, such as precision, recall, F1 score, error detection rate, total training time, total test time, average training time, Type I error, and Type II error were calculated and displayed to the user. Besides, the user is offered a voting option for deciding the algorithm to use. Selected graphical user interfaces of the application are provided in Fig. 6. The application enables users to select the machine learning model for a given set of data and then compares the output using graphical elements. For developing and testing purposes, AML Workshop dataset from Microsoft [42] is used in the MLL module.

### Hybrid Digital Twin

The above-described platform is enhanced by DT, which contains two related models: Data-Driven and Physics-Driven (first-order principal models); thus a Hybrid DT is generated.

In the context of hybrid Digital Twins, Physics-Driven models can be beneficial over the Data-Driven ones in many aspects, including but not limited to:

- **Generating synthetic data in case of “data-poor” cases:** A typical example is training a ML pipeline for predictive maintenance. Meanwhile, very often, when a machine is new, it does not have historical sensor data that can be used to train a data-driven approach. When carefully designed, the virtual physics-based twin can generate the needed supervised training dataset.
- **Quality control of data-driven Digital Twin:** When operating a critical infrastructure or asset, it is seen as a risky approach to fully rely on data-driven approaches in taking real-time decisions. To mitigate such risks, it is possible to build a control pipeline in which the physics-based model will be a controller to the data-driven predictor. A broker needs to be designed to integrate the two approaches in a seamless way.

On the other hand, data-driven models can be used to continuously calibrate physics-based models. In fact, machine degradation, wearing of parts, environment, and other factors impact the overall process performance over time. The state of the practice is that an operator will manually recalibrate the control system when a deviation is identified. Such manual operation can be replaced by setting a data-driven model that will identify and calibrate critical process variables that will be fed into a physics-based model that will optimise the process’s control system.

An example of first-order model tools we have is a DC motor, which is a commonly used component. The motor is made of two coupled electrical and mechanical models based on the governing equations from Newton’s second law ( $v = R \cdot i + L \cdot di/dt + ve$ ) and Kirchhoff’s voltage law ( $Te = TL + B \cdot \omega + JL \cdot d\omega/dt$ ) where  $v$  is the voltage [V],  $i$  is the current [A],  $ve$  is the back electromotive force

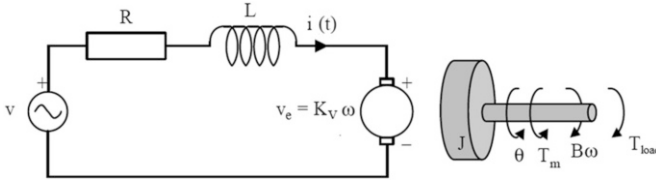


Fig. 11 Physics-based schematic of a DC motor (<http://pubs.sciepub.com/ajme/4/7/27/>)

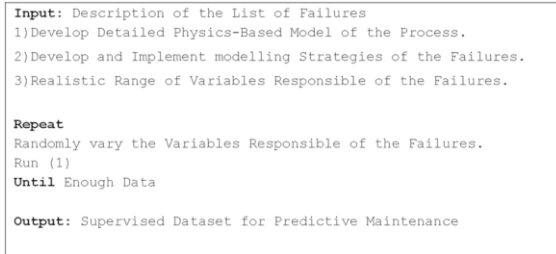


Fig. 12 Synthetic data generation for Predictive Maintenance Pipeline

[V],  $\omega$  is the angular velocity [rad/s], TL is the Load torque [Nm], JL is Load inertia + Rotor inertia [kg·m<sup>2</sup>],  $v_e = K_v \cdot \omega$ ,  $T_e = K_t \cdot i$ ,  $K_t$  is Torque constant [Nm/A],  $K_v$  is the back EMF constant [V/(rad/s)],  $R$  is the Phase Resistance [Ohm],  $L$  is the Phase Inductance [H],  $J$  is the Rotor Inertia [kg·m<sup>2</sup>], and  $B$  is the Rotor Friction [Nm/(rad/s)] (Fig. 11).

A hybrid Digital Twin for predictive maintenance of a DC motor is built using the principles given above and the architecture given in Fig. 12. For the physics-based model, the MATLAB environment has been used to model the physical elements of the electric DC motor and related electric and mechanical components. Real data has been obtained by means of real sensor data collected, but they are not with enough failure cases to train an ML model. On the DC motor, the following sensors have been installed (current, voltage, temperature, and vibration). The physics-based model is being calibrated by comparing the measured data with the MATLAB predicted one. Results show a qualitative similarity between the synthetic and the sensors data. The models will later be added to the knowledge base by integrating the cognitive elements into the model.

### Cognitive Digital Twin

In the Steel Pipe Welding pilot case, CT will introduce improved decision-making by integrating human knowledge into the decision-making process. The anomalies, alarms, and early warnings of machine and system problems will be tackled by CT, and the decision-making process will emulate the experienced human operator with the embedded knowledge base. CT will augment expert knowledge for unforeseeable cases on HT and DT. The human operator’s knowledge is reflected to process knowledge and physics-based models with parametric values as well as

thresholds and causality relations. Expert knowledge on the causes of breakdowns is collected with the series of problematic operations and the initial causes that trigger the successive reactions.

Cognition will be integrated to support Cognitive Digital Twins for learning, understanding, and planning, including the use of domain and human knowledge by making use of the ML algorithms, ontologies, and Knowledge Graphs (KG) to capture background knowledge, entities, and their relationships. CT will bring life cycle optimisation by reacting to early warnings and suggesting optimised predictive actions to improve operational performance by optimising operational parameters and enhance the utilisation of computing infrastructure and energy usage.

#### ***6.4 Digital Twin Visualisation and Control***

This component contains dashboards suitable for sensor data, error detection, and transfer of regular information obtained from data processing to the real-time status monitoring system, and development of end-user applications.

Three.js, an open-source JavaScript library, was used to develop animated or non-animated 3D applications that can be opened in the web browser using WebGL. Three.js is supported by all WebGL-supported web browsers. In addition to Three.js for visualisation of the Digital Twin elements, Solidworks [43] is used for 3D visualisation.

For the web interface, the JSON data received with JavaScript have been parsed and then transferred to PHP pages. In this communication, the post method has been used in the requests sent with JavaScript. With the help of PHP, the information is placed in HTML objects. Grafana [44] is used in the process of placing graphics within the card object. Dynamic graphics created on Grafana are placed on cards in iframe tags.

Sensor data on temperature, vibration, pressure, current, and oil temperature are given in dashboards. Alarm and status information are provided as they occur. Besides sensor, alarm, and status values, real-time monitoring of parameters like pipe diameter, pitch angle, belt width, production speed, motor speed (RPM), energy consumption, instant output power, pipe diameter, and wall thickness are visualised in the dashboards.

### **7 Conclusions**

This chapter has presented the Digital Twin Pipeline Framework of the COGNITWIN project that supports Hybrid and Cognitive Digital Twins, through four Big Data and AI pipeline steps. The approach has been demonstrated with a Digital Twin pipeline example for Spiral Welded Steel Pipe Machinery maintenance. The

components used in this pipeline have also been mapped into the different technical areas of the BDV Reference Model.

The COGNITWIN project has further five use case pilots which include also other types of sensors, in particular for image and video analytics with RGB and infrared cameras, and support for image analytics through deep learning, including AI@Edge support through FPGA hardware. This includes also high temperature processes for aluminium, silicon, and steel production and Digital Twin pipeline support including analytics for nonlinear model predictive control combined with machine learning.

Synergies between and combinations of different elements for hybrid Digital Twin are now being further enhanced through the use of orchestration technologies like StreamPipes and Digital Twin access APIs like AAS.

The COGNITWIN project is now proceeding with further Hybrid Digital Twins in all of the six pilot cases, extending this also with cognitive element for self-learning and control, for the establishment of a Toolbox with relevant and reusable tools within each of the four pipeline steps. This will be further applied and evaluated for the six pilot use cases of the project, including the steel welding pilot described in this chapter.

**Acknowledgements** The COGNITWIN project has received funding from the European Union's Horizon 2020 research and innovation programme under GA No.870130. The platform developed by TEKNOPAR [45] is partially funded by the COGNITWIN project and has been validated by the NOKSEL pilot of the project. We thank the COGNITWIN consortium partners for fruitful discussions related to the Digital Twin pipeline architecture and the use case presented in this chapter.

## References

1. Abburu, S., Berre, J. A., Jacoby, M., Roman, D., Stojanovic, L., & Stojanovic, N. (2020, November). Cognitive digital twins for the process industry. In *Cognitive technologies, the twelfth international conference on advanced cognitive technologies and applications (COGNITIVE 2020)*.
2. Albayrak, Ö., & Unal, P. (2020). Smart Steel pipe production plant via cognitive digital twins: a case study on digitalization of spiral welded pipe machinery. In *The proceedings of the ESTEP workshop on impact and opportunities of artificial intelligence in the steel industry*.
3. Jacoby, M., & Usländer, T. (2020). Digital twin and internet of things—Current standards landscape. *Applied Sciences*, 10, 6519. Retrieved from <https://www.mdpi.com/2076-3417/10/18/6519>
4. Eirinakis P., Kalaboukas, K., Lounis, S., Mortos, I. Rozanec, J. M., Stojanovic, N., & Zois, G. (2020). Enhancing cognition for digital twins. In *IEEE International conference on engineering, technology and innovation (ICE/ITMC), Cardiff, United Kingdom, 2020* (pp. 1–7). <https://doi.org/10.1109/ICE/ITMC49519.2020.9198492>.
5. Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: enabling technologies, challenges and open research. *IEEE Access*, 8, 108952–108971. <https://doi.org/10.1109/ACCESS.2020.2998358>
6. Abburu, S., Berre, J. A., Jacoby, M., Roman, D., Stojanovic, L., & Stojanovic, N. (2020, June 15–17). COGNITWIN—hybrid and cognitive digital twins for the process industry.

- In *Proceedings of the 2020 IEEE international conference on engineering, technology and innovation (ICE/ITMC)*, Cardiff, UK (pp. 1–8).
7. Berre, A. J., Tsalgaidou, A., Francalanci, C., Ivanov, T., Lobo, T. P., Saiz, R. R., Novalija, I., & Grobelnik, M. (2021). Big data and AI pipeline framework—Technology analysis from a benchmarking perspective. In *TABDV book*. Springer Open Access.
  8. Krupitzer, C., Wagenhals, T., Züfle, M., Lesch, V., Schäfer, D., Mozaffarin, A., Edinger, J., Becker, C., & Kounev, S. (2020). A survey on predictive maintenance for Industry 4.0. *arXiv*, preprint arXiv:2002.08224.
  9. Wu, S. J., Gebrael, N., Lawley, M. A., & Yih, Y. (2007). A neural network integrated decision support system for condition-based optimal predictive maintenance policy. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 37(2), 226–236. <https://doi.org/10.1109/TSMCA.2006.886368>
  10. Baidya, R., & Ghosh, S. K. (2015). Model for a predictive maintenance system effectiveness using the analytical hierarchy process as analytical tool. *IFAC-PapersOnLine*, 48(3), 1463–1468.
  11. Deloux, E., Castanier, B., & Bérenguer, C. (2009). Predictive maintenance policy for a gradually deteriorating system subject to stress. *Reliability Engineering & System Safety*, 94(2), 418–431.
  12. Dayarathna, M., & Perera, S. (2018). Recent advancements in event processing. *ACM Computing Surveys (CSUR)*, 51(2), 1–36.
  13. Fülöp, L. J., Beszédes, Á., Tóth, G., Demeter, H., Vidács, L., & Farkas, L. (2012, September). Predictive complex event processing: a conceptual framework for combining complex event processing and predictive analytics. In *Proceedings of the fifth Balkan conference in informatics* (pp. 26–31).
  14. Sahal, R., Breslin, J. G., & Ali, M. I. (2020). Big data and stream processing platforms for Industry 4.0 requirements mapping for a predictive maintenance use case. *Journal of Manufacturing Systems*, 54, 138–151.
  15. Calabrese, M., Cimmino, M., Fiume, F., Manfrin, M., Romeo, L., Ceccacci, S., Paolanti, M., Toscano, G., Ciandrini, G., Carrotta, A., Frontoni, E., Kapetis, D., & Mengoni, M. (2020). SOPHIA: An event-based IoT and machine learning architecture for predictive maintenance in Industry 4.0. *Information*, 11(4), 202.
  16. Aivaliotis, P., Georgoulis, K., & Chryssolouris, G. (2019). The use of digital twin for predictive maintenance in manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(11), 1067–1080.
  17. Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), 1483–1510.
  18. Paolanti, M., Romeo, L., Felicetti, A., Mancini, A., Frontoni, E., & Loncarski, J. (2018, July). Machine learning approach for predictive maintenance in industry 4.0. In *2018 14th IEEE/ASME international conference on mechatronic and embedded systems and applications (MESA)* (pp. 1–6). IEEE.
  19. Dalzochio, J., Kunst, R., Pignaton, E., Binotto, A., Sanyal, S., Favilla, J., & Barbosa, J. (2020). Machine learning and reasoning for predictive maintenance in Industry 4.0: Current status and challenges. *Computers in Industry*, 123, 103298.
  20. Li, Z., Wang, Y., & Wang, K. S. (2017). Intelligent predictive maintenance for fault diagnosis and prognosis in machine centers: Industry 4.0 scenario. *Advances in Manufacturing*, 5(4), 377–387.
  21. Carvalho, T. P., Soares, F. A., Vita, R., Francisco, R. D. P., Basto, J. P., & Alcalá, S. G. (2019). A systematic literature review of machine learning methods applied to predictive maintenance. *Computers & Industrial Engineering*, 137, 106024.
  22. Schmidt, B., Wang, L., & Galar, D. (2017). Semantic framework for predictive maintenance in a cloud environment. *Procedia CIRP*, 62, 583–588.
  23. Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Sittón, I., & Corchado, J. M. (2019, May). A predictive maintenance model using recurrent neural networks. In *International*

- workshop on soft computing models in industrial and environmental applications* (pp. 261–270). Springer.
24. Kanawaday, A., & Sane, A. (2017, November). Machine learning for predictive maintenance of industrial machines using IoT sensor data. In *2017 8th IEEE international conference on software engineering and service science (ICSESS)* (pp. 87–90). IEEE.
  25. Yuan, Y., Ma, G., Cheng, C., Zhou, B., Zhao, H., Zhang, H. T., & Ding, H. (2018). Artificial intelligent diagnosis and monitoring in manufacturing. *arXiv*, preprint arXiv:1901.02057. <https://doi.org/10.1093/nsr/nwz190>
  26. De Vita, F., Bruneo, D., & Das, S. K. (2020, April). A novel data collection framework for telemetry and anomaly detection in industrial IoT systems. In *2020 IEEE/ACM fifth international conference on internet-of-things design and implementation (IoTDI)* (pp. 245–251). IEEE.
  27. NOKSEL. <http://www.noksel.com.tr/index.php/en/>
  28. Apache Kafka. <https://kafka.apache.org/>
  29. MQTT Broker. <https://mqtt.org/>
  30. Open Platform Communications, OPC. <https://opcfoundation.org/about/what-is-opc/>
  31. BDVA, Big Data Value Association. <https://www.bdva.eu>
  32. 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.
  33. Docker. <https://www.docker.com/>
  34. Apache Spark. <https://spark.apache.org/>
  35. Unal, P. (2019). Reference architectures and standards for the internet of things and big data in smart manufacturing. In *2019 7th international conference on future internet of things and cloud (FiCloud)* (pp. 243–250). IEEE.
  36. Apache Cassandra. <https://cassandra.apache.org/>
  37. Python. <https://www.python.org/>
  38. PostgreSQL. <https://www.postgresql.org/>
  39. OPC UA. <https://opcfoundation.org/about/opc-technologies/opc-ua>.
  40. Apache StreamPipes. <https://streampipes.apache.org/>
  41. Scikit-Learn. <https://scikit-learn.org/>
  42. AML Data Set, <https://github.com/microsoft/AMLWorkshop/tree/master/Data>
  43. Solidworks. <https://www.solidworks.com/tr>
  44. Grafana. <https://grafana.com/>
  45. TEKNOPAR. <https://TEKNOPAR.com.tr>

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

