Ref. Ares(2020)4619539 - 04/09/2020

IoTwins
BIG DATA OPTIMIZE INDUSTRY AND SERVICES

Grant Agreement N°857191

Distributed Digital Twins for industrial SMEs: a big-data platform

Deliverable 4.1 – Enhanced data collection and integration for the manufacturing test-beds

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement Nº 857191

1. Document Identification

Project	IoTwins
Project Full Title	Distributed Digital Twins for industrial SMEs: a big-data platform
Project Number	857191
Starting Date	September 1st, 2019
Duration	3 years
H2020 Programme	H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT)
Topic	ICT-11-2018-2019 - HPC and Big Data enabled Large-scale Test-beds and Applications
Call for proposal	H2020-ICT-2018-3
Type of Action	IA-Innovation Action
Website	iotwins.eu
Work Package	WP4
WP Leader	GCL
Responsible Partner(s)	GCL (with the contribution of: FILL, BRI, BSC, ETXE)
Contributors	Edoardo Massano (GCL), Verena Stanzl (FILL), Gaetano Ciaravella (BRI), Javier Diaz (ETXE), Filippo Mantovani (BSC)
Reviewers	Claudio Domenico Arlandini (CINECA), Jean-Christian Ahouangonou (ESI)
File Name	D 4.1 – Enhanced data collection and integration for the manufacturing test-beds
Contractual delivery date	M12 - 31 August 2020
Actual delivery date	M13 - 04 September 2020
Version	1
Status	Final
Dissemination level	Public
Author	Tommy Langers (GCL)
Contact details of the coordinator	Francesco Millo, francesco.millo@bonfiglioli.com

Table of Contents

2. Introduction

The main goal of work package (WP) 4 is the definition and implementation of four large-scale industrial testbeds in the manufacturing sector. They allow within WP2 and WP3 the definition of the computing architecture that will enable a digital twin platform to collect the data. The available IT infrastructures were used to develop components linked to the deliverables D2.1 (Architecture and Technical/User Requirements (I)) and D3.1 (Requirements from large-scale industrial production and facility digital twins) and the milestone MS2 (IoTwins Platform and services – Requirements form large-scale industrial production, finalized platform specifications and facility digital twins available).

This deliverable reports on the work performed within the four large-scale industrial testbeds between M1 and M12 of the project.

2.1. Work package objectives

On-line data and off-line historical data from industrial plants and machinery will be used to generate an upstream flow of data to feed the developed components. To realize the required upstream of data, industrial control technology and IT solutions are implemented to the manufacturing testbeds.

By collaborating with WP2 and WP3, the concept of the first digital twin platform was defined by sharing the requested requirements. The architecture and the data flow are designed that every testbed can access the platform in a homogenous way.

Deliverable D4.1 will focus on:

- enhanced data collection
- data fusion
- distributed twin creation
- validation/generalization of the results.

2.2. Methodology

The testbeds exhibit several elements of heterogeneity in their IT infrastructure, also because they involve industrial processes implemented before the IoTwins project. In addition, each one is subject to different constraints/requirements due to the different industrial branches. In strict collaboration with WP2 and WP3, this work package has tried to adapt a general architecture and methodology for all the testbeds with tailormade specializations for the different cases.

The following sub-sections describe the generalized methodology to achieve the requested results.

2.2.1. Identification of the use case

At the first stage of the project the activity was to figure out the best scenario of a connected machine to identify the use case for the pilot. The main goal was to provide a huge quantity of data to create a model to be analysed in the following project phases.

2.2.2. Technical preparation

To allow industrial plants/machineries to upstream production data to IT services, it is often necessary to perform hardware- and/or software-sided upgrades. By collaborating with e.g. suppliers, clients, the machines can be enhanced technically to fulfil the required features. Also, in some cases, laboratory setups are required.

2.2.3. Identification of sensors

Based on the specific testbeds, the requirements for the monitored events related to the machineries are slightly different. Such an event can be e.g., a breakdown to be avoided before it happens. A set of sensors monitoring the environment in each testbed has been identified.

2.2.4. Data processing

The processing of the data is executed in the dedicated infrastructures and/or applications defined in WP2 and WP3 for each testbed.

2.2.5. Digital twin

By defining the expectations of the usage of a digital twin for each testbed, the correct development of the first platform version is realized. In this subsection, testing and describing the behaviour of the digital twin is foreseen. On top of it, a more general concept of the digital twin will allow a vision of the platform in the near future.

2.2.6. Validation

The following points can validate the individual tasks:

- investigation of the correct collecting of data
- security aspects
- proper usage/upstream of data.

3. Test-bed 1: Wind-farm predictive maintenance system, BRI

3.1. Abbreviations

Within this section, the following abbreviations are commonly used:

FEM – Finite-element method

MOR – Model-order reduction

MBS – Multibody system/simulation

PGD – Proper generalized decomposition

ROM – Reduced-order model

WTG – Wind turbine generator.

3.2. Identification of the use case

In test-bed 1 we have identified as use case a wind turbine generator (WTG) located at the end user's wind park, where several types of wind turbine are installed. Some of these are exposed to increased turbulence and sudden changes in wind direction. These kinds of aleatory loads represent a source of possible additional structural, mechanical and electrical stresses on WTGs.

In addition to the uncertain and "unpredictable" weather conditions, the fatigue and natural ageing processes on mechanical and electrical parts of a WTG represent a source of possible faults.

A WTG works thanks to its rotor, a common term for the three blades and hub, harvests a part of the wind's kinetic energy, thereby causing the rotor's main shaft to turn and generate a certain torque. The main shaft is, typically via a gearbox, connected to a generator which then converts the mechanical energy into electrical energy.

To achieve the best utilization of the wind's kinetic energy, the aerodynamic angle of attack of the blades is kept at its optimum value.

The first generation of modern wind turbines were stall-controlled, the blades being mounted to the hub at a fixed angle. The aerodynamic design of the blades ensured that, when the wind speed became too high for the turbine to operate safely, turbulence was created on the backside of the blades, thus preventing the kinetic energy of the wind acting on the rotor rotation alone.

The rotor on newer generations of wind turbines are pitch-controlled, where the position of the blades, and thereby the angle of attack is not fixed, but rather controlled by either an electrical or hydraulic system as a function of the power output, wind speed, structural loads etc. Such pitch control of the blades allows better energy yield at lower wind speeds and especially at higher wind speeds compared to stall-controlled turbines.

A third scheme is called active stall, which is the technology used by the target wind turbine in testbed 1. This design was commonly used in the transition from passive stall-controlled turbines to fully pitch-controlled turbines. Technically, active stall resembles pitch control since, as both wind turbine types involve pitching the blades to get an optimal aerodynamic torque at low wind speeds. The main difference between active-stall-controlled and modern pitch-controlled turbines occurs at the point when the wind turbine is operating at rated power and the wind speed approaches the point of overloading the generator. Where rated, or maximum, power is the active power that the turbine will produce when the wind is in the ideal range for operation.

The active-stall-controlled wind turbine will react by pitching its blades in the opposite direction to that which a pitch-controlled wind turbine does. Hereby, it will increase the angle of attack of the rotor blades to make the blades go into a stall. Compared to a pitch-controlled turbine, an active-stall-controlled turbine thus wastes the excess energy generated by the wind.

Common to all three concepts above, is an operating yaw system. While the pitch system controls the position of the individual blade, the yaw system controls the position of the rotor. If the yaw system does not keep the rotor facing the wind, then the area of the rotor closest to the wind direction will be subject to higher forces compared to the rest of the rotor and thus the wind turbine will be subject to fatigue loads as well as potentially hazardous loads.

For exploring the causes that can generate failures in WTGs, we focused our investigation mainly on the nacelle (Figure 1).

Figure 1: Nacelle

Figure 2 depicts a typical wind turbine cost distribution[1]. However, what is not visualized and represented is the cost of failure coming from several parts. An example can be provided from the yaw system that represents only 1,25% of the entire cost of a wind turbine, but whose failure and consequent stop can generate higher costs.

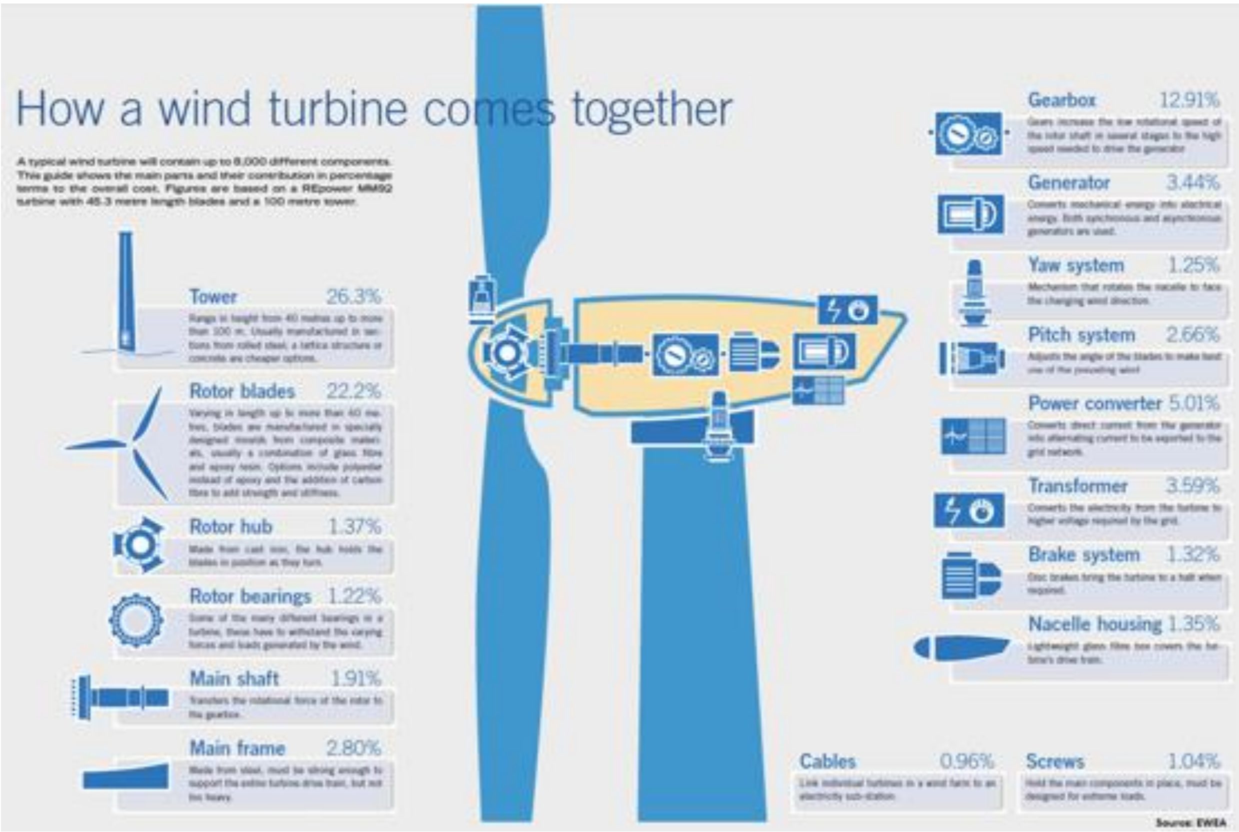


Figure 2: Wind turbine cost distribution

[1] Wind Directions, January/February 2007

With corrective actions (after a failure occurs) the estimated costs of intervention/repair vary from 60k€ to 200k€ per turbine; this is in addition to the cost of production loss during the intervention (2k € to 4k€)[2].

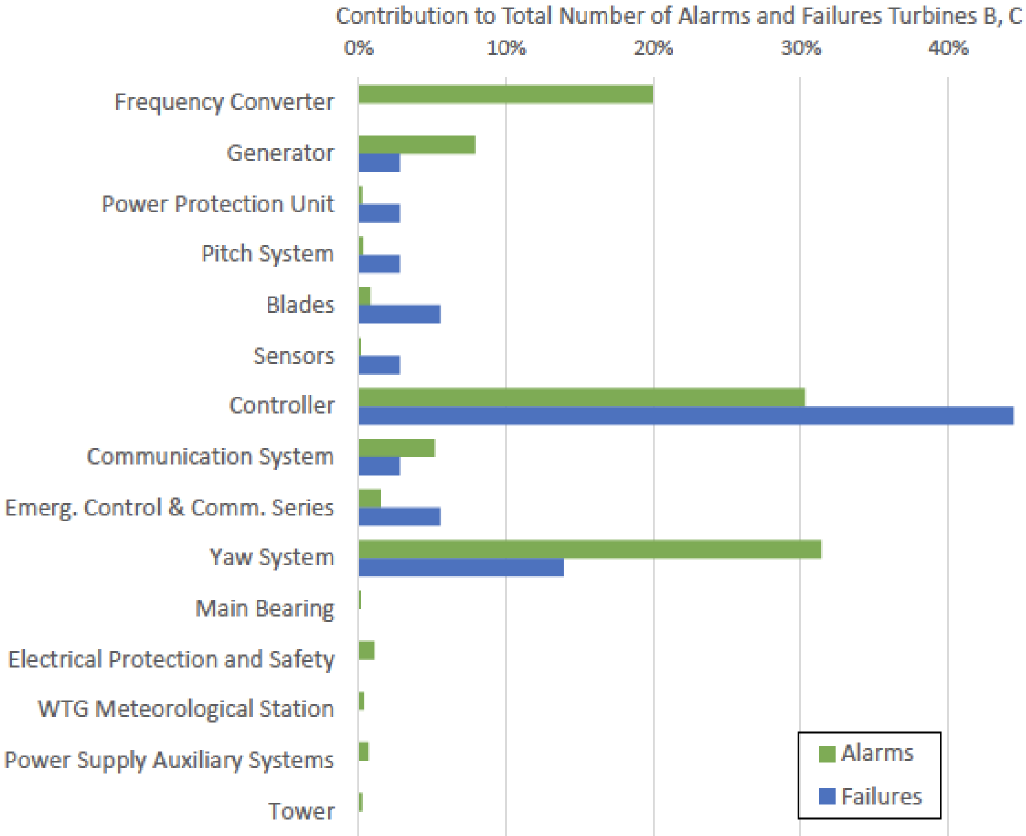
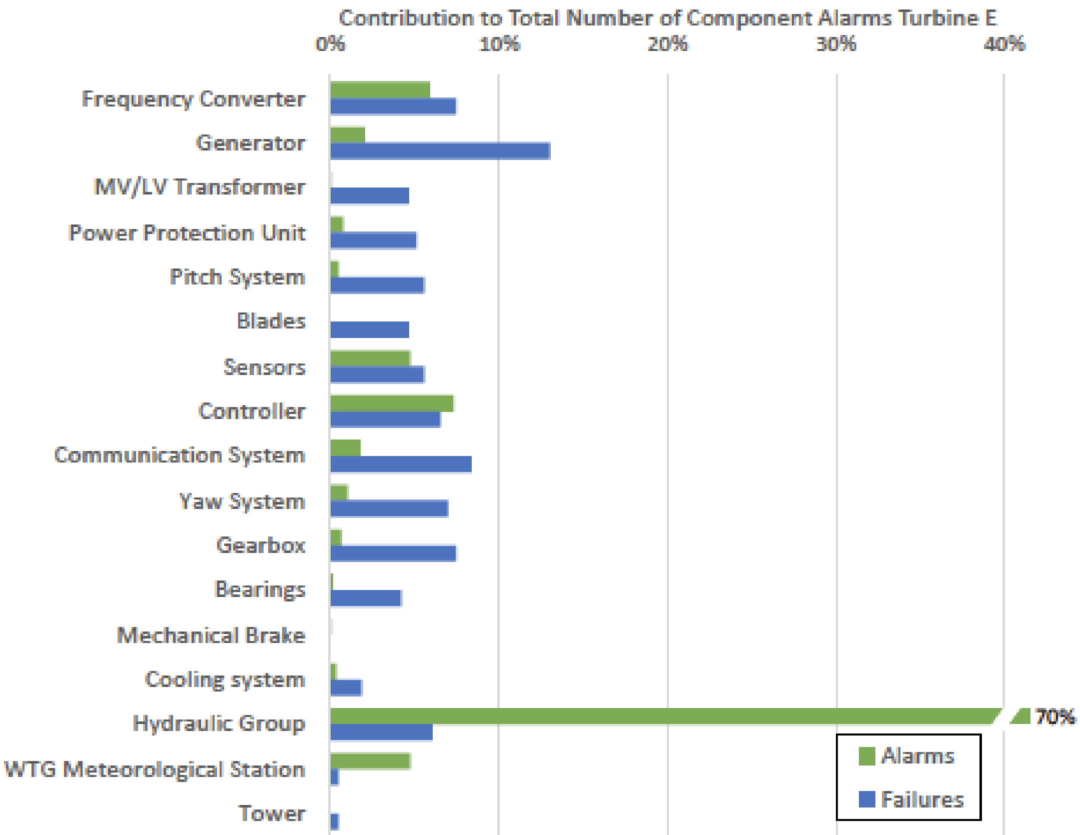


Figure 3: Contribution to total number of alarms and failures turbines B, C (B and C are WTGs with a rated capacity of 2MW)[3]

[2] *Wind Turbine Failures - Tackling current Problems in Failure Data Analysis: Journal of Physics: Conference Series 753 (2016) 072027*

[3] Journal of Physics: Conference Series 753 (2016) 072027; *Wind Turbine Failures - Tackling current Problems in Failure Data Analysis*



(d) Turbine Type E

Figure 4: Contribution to total number of alarms and failures turbine E (E is a WTG with a rated capacity of 2MW)

Avg. yearly number of wind farms	230
Mean yearly installed capacity (MW)	5818
Avg. number of failure events per year	2280
Avg. yearly number of WTs considered	over 4300
Containing:	
Avg. yearly number of WTs under 1 MW	2130
Avg. yearly number of WTs equal or over 1 MW	2270
Number of direct drive turbines	215

Figure 5: Additional statistics related to wind farms

The figures (Figure 3, Figure 4, Figure 5) above represent, in several wind turbine generators, a description of root causes and also the correlation on the alarms generated, highlighting a big difference between the number of alarms and number of real failures related to the type of alarm.

An approach where data acquired with high frequency, Machine Learning and Hybrid Twins can represent a significant step forward in the direction of more pragmatic and real detection and prediction of failures.

3.3. Technical preparation

The WTG (Figure 6) identified is a wind turbine type that can generate 1.3MW power and which has been on the market for 20 years. The end user highlighted that this type of wind turbine often requires mechanical and electrical part substitutions due to prolonged usage and age.

Figure 6: Picture of a WTG

The technical characteristics of the WTG are the following [4]:

3.3.1. Power

Rated power: 1,300.0 kW

Flexible power ratings: -

Cut-in wind speed: 3.0 m/s

Rated wind speed: 15.0 m/s

Cut-out wind speed: 25.0 m/s

Survival wind speed: 55.0 m/s

Wind zone (DIBt): -

Wind class (IEC): -

3.3.2. Rotor

Diameter: 62.0 m

[4] https://en.wind-turbine-models.com/turbines/59-an-bonus-1300-62#datasheet

Swept area: 3,019.0 m²

Number of blades: 3

Rotor speed, max: 19.0 U/min

Tipspeed: 62 m/s

Type: LM 29/B30

Material: GFK

Manufacturer: LM Glasfieber/Bonus

Power density 1: 430.6 W/m²

Power density 2: 2.3 m²/kW

Gear box

Type: Spur/planetary

Stages: 3.0

Ratio: 1:79

Manufacturer: Flender

Generator

Type: Asynchronous

Number: 1.0

Speed, max: 1,500.0 U/min

Voltage: 690.0 V

Grid connection: Thyristor

Grid frequency: 50.0 Hz

Manufacturer: ABB

3.3.3. Tower

Hub height: 68/80/90 m

Type: Tubular steel

Shape: Conical

Corrosion protection: Coated

Manufacturer: Roug/KGW/SSC

3.3.4. Weight

Single blade: -

Hub: -

Rotor: 30.0 t

Nacelle: 50.0 t

Tower, max: 198.0 t

Total weight: 280.0 t

3.3.5. Power curve

The power curve of the wind turbine represents the generated power as a function of the wind speed and is a key metric for the operational performance. A representative power curve for the target wind turbine is shown below[5].

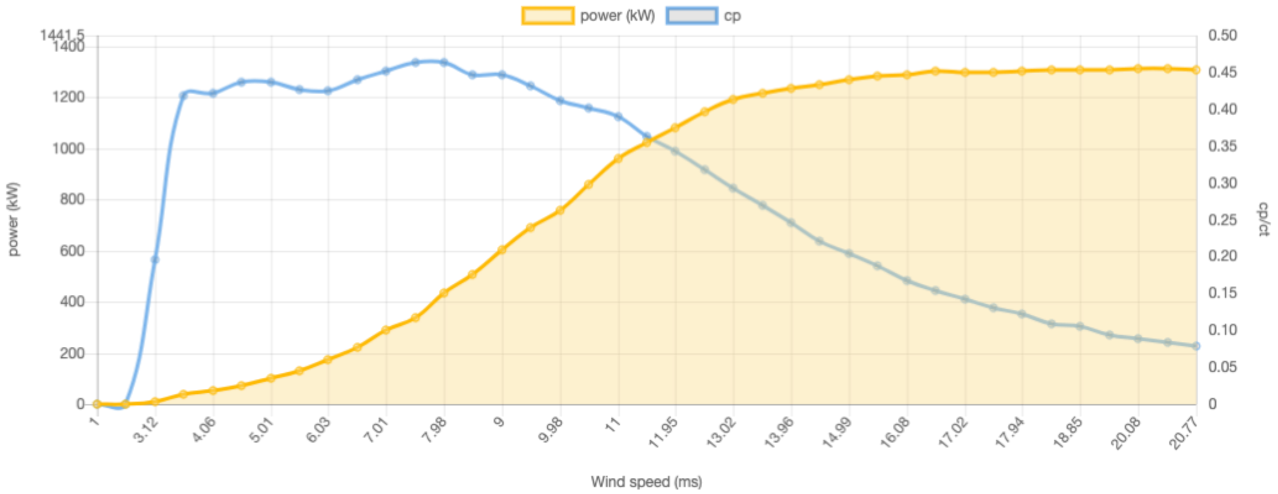


Figure 7: Power curve of the wind turbine

The WTG is controlled and is already able to transmit data and failures via the turbine control system.

3.4. Identification of the sensors

The aim is to merge two systems through an Edge Computer being capable of gathering data from these systems, both the turbine control system and the purpose-built multi-sensor platform system.

[5] https://en.wind-turbine-models.com/turbines/59-an-bonus-1300-62#powercurve

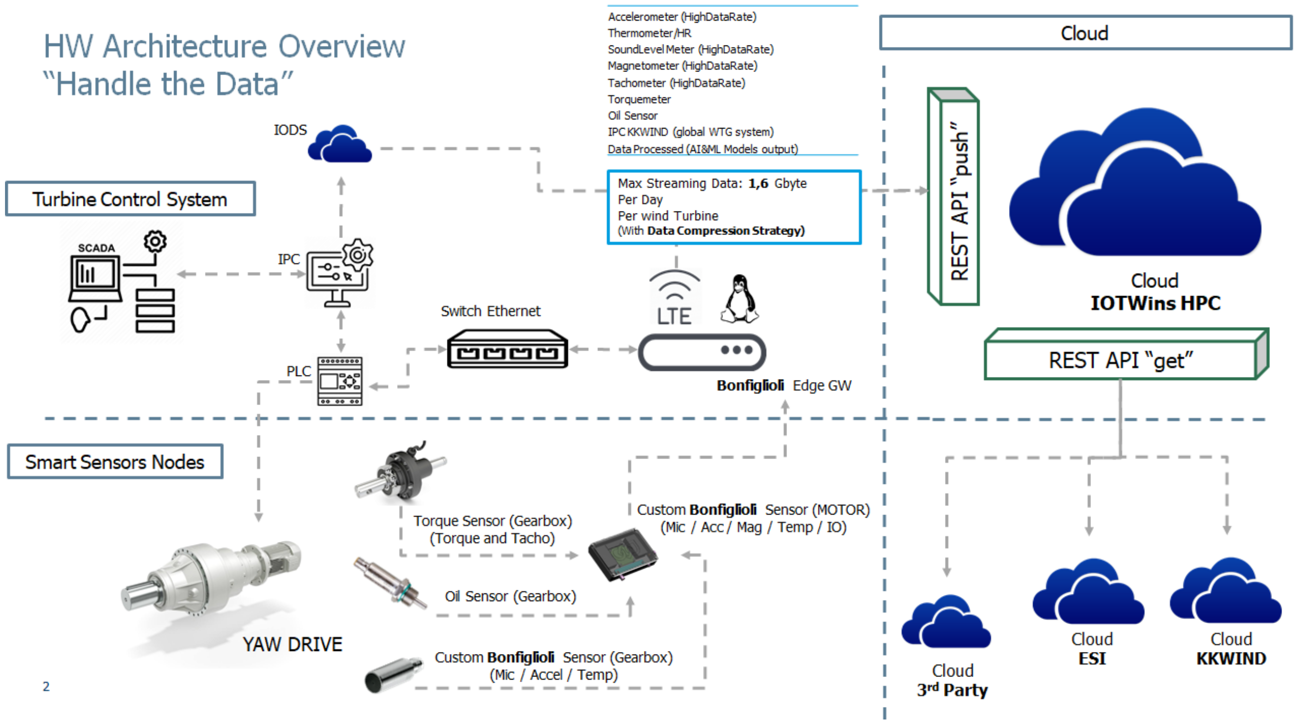


Figure 8: Hardware architecture overview

The two systems consist of smart sensors nodes from one side and on the other side the turbine control system plus the SCADA of the WTG.

The data stream that will be generated, based on team experience and know-how, will be based on both sensors applied on gearmotors and data coming from turbine control systems and possibly ecosystems.

The data types of the gathered sensor data are the following:

- Accelerometer (RAW)
- Thermometer
- HR (relative Humidity)
- Sound Level Meter (RAW)
- Magnetometer (RAW)
- Speed meter
- Speed meter (RAW)
- Torque meter
- Oil GB Sensor (Full version)
- Timestamp (NTP)
- Data source, WTG control system
- Live data for Edge Twin (Current/default data configuration)
- Live data for Edge Twin (Extensive data configuration)
- Historical data for Edge Twin (Extensive data configuration)
- Weather forecast (OPTION)
- Energy prices (OPTION)
- Meteorology mast (OPTION)

The following table summarises both the data type, size and sampling frequency.

Table 1: Gathered sensor data of the yaw motor

Yaw Motor Data Source	Quantity/Format	Sample/s	MB/day produced	MB/day stored in the IoT Layer	Notes
Accelerometer (RAW)	12 (Float32)	100 000	1098.63	109.8	10KHz for 10sec of the signal
Thermometer (HR)	8 (Float16)	1	1.31	1.31	1Hz for the continue signal
Sound Level Meter (RAW)	4 (Float32)	100 000	366.2	36.62	10KHz for 10sec of the signal
Magnetometer (RAW)	12 (Float32)	100 000	1098.63	109.8	10KHz for 10sec of the signal
Speed meter	4 (Float16)	4	5.2	5.2	4Hz for the continue signal
Speed meter (RAW)	4 (Float32)	15 000	54.9	5.49	1.5KHz for 10sec of the signal
Torque meter	4 (Float16)	4	5.2	5.2	4Hz for the continue signal
Oil GB Sensor (Full version)	60 (Float16)	4	>1	>1	1Hz for the continue signal
Timestamp (NTP)	8 (Integer64)	1000	>1	>1	100Hz for the continue signal

Table 2: Gathered sensor data of the yaw gearbox

Yaw Gearbox Data Source	Quantity/Format	Sample/s	MB/day produced	MB/day stored in the IoT Layer	Notes
Accelerometer (RAW)	12 (Float32)	100 000	1098.63	109.8	10KHz for 10sec of the signal
Thermometer HR	8 (Float16)	1	1.31	1.31	1Hz for the continue signal

Yaw Gearbox Data Source	Quantity/Format	Sample/s	MB/day produced	MB/day stored in the IoT Layer	Notes
Sound Level Meter (RAW)	4 (Float32)	100 000	366.2	36.62	10KHz for 10sec of the signal
Timestamp (NTP)	8 (Integer64)	1000	>1	>1	100Hz for the continue signal

Table 3: Specifications of the data sources

<table>
<tr><th>Data source, WTG control system</th><th>Quantity/Format</th><th>Samples/s</th><th>Data/day</th><th>MB/day, stored in the IoT Layer</th><th>Notes</th></tr>
<tr><td rowspan="1">Live data for Edge Twin (Current/default data configuration)</td><td>- 1400 Variables

(IOs, filters, state variables, summations, derived calculations etc.)

- 180 Events

(Info, warnings and alarms)

- 630 Parameters

(Monitoring and control thresholds, can be changed runtime)</td><td>Individual sample rate of each signal.

(max sample rate is 100Hz.)</td><td>Up to ~10 GB</td><td rowspan="4">Same amount as generated.</td><td rowspan="4"></td></tr>
<tr><td rowspan="3">Live data for Edge Twin (Extensive data configuration)</td><td>Same quantity as above, specified in more detail below…</td><td rowspan="3">All signals are logged by 100Hz.

(max sample rate is 100Hz.)</td><td>Up to ~60 GB</td></tr>
<tr><td>320 boolean variables (1 bit).</td><td>~330 MB</td></tr>
<tr><td>1080 float and integer variables (32 bit)</td><td>~35600 MB</td></tr>
</table>

	180 boolean events (1 bit).		~190 MB		
	8 boolean parameters (1 bit).		~8 MB		
	622 float and integer parameters (32 bit)		~20500 MB		
Historical data for Edge Twin (Extensive data configuration)	Approx. 2100 data points incl. variable names and timestamp.	All signals are logged by 100Hz. (max sample rate is 100Hz.)	Up to ~**140 GB**		All live data generated by the WTG is stored as historical data and can be used for training of Hybrid Twin.
Weather forecast (OPTION)	~ 460 bytes pr. forecast	1 Hz (circa)	~**40 MB**	-	
Energy prices (OPTION)	**Unknown**	1 Hz (circa)	*~40 MB*	-	
Meteorology mast (OPTION)	3 additional anemometers 3 additional wind vanes (both 32 bit)	Logged by 100Hz	~**200 MB**	-	

On the Edge Twin Layer, it was possible to define the type of analyses to be performed based on the data type as shown in the table below:

Table 4: Type of analyses for the edge layer

Data source	Computation type	Data Produced (size, rate and type)	Expected computational power needed	Notes
All data gathered into the system	• Filtering • Aggregating • Decisions to be taken • Simulations to be performed • Data produced • Big Data Analytics • Monitoring dashboard • ML applications • Monitoring dashboard	About 1 GB	About 2.3 DMIPS/MHz@1000MH	
All data gathered into the system	• Lookup table search • Floating point operation	About 100 MB	About 2.3 DMIPS/MHz@1000MH	• Predictive maintenance • Energy production • WTG extended life
WTG control parameters	• Buffering	About 1.5 GB	About 2.3 DMIPS/MHz@1000MH	• ML algorithms • Models

Therefore, the type of analyses defined both the type of EDGE Computer required to achieve the performances identified by the team and the differences between the current edge hardware used in other wind turbine applications.

Table 5: Specifications of the edge devices

Module name/type	Computing resources	RAM	Storage Capacity	Networking Capabilities	Notes
Edge GW (MPU)	2.3 DMIPS/MHz @1000MHz (2 core)	2 Gbyte	32 Gbyte SSD	2x Ethernet 3G/4G WiFi	Cost Optimized Current solution

Module name/type	Computing resources	RAM	Storage Capacity	Networking Capabilities	Notes
Edge GW (MPU)	6 DMIPS/MHz @1500MHz (4 core) 64bit	4 Gbyte	128 Gbyte SSD	2x Ethernet 3G/4G WiFi	Performance Optimized based on WP4.1 Testbeds Partners' needs

3.5. Data processing

The Edge GW will collect data from the sensorized yaw drives while the turbine control system will store historical data in a cloud-based time series database (IODS). Both the Edge GW and IODS will push historical data to the IoTwins HPC computing systems, a time server will be used to ensure timestamp of data sets are synchronized.

The Edge GW will furthermore have full access to the turbine data acquisition system, meaning all variables (IOs, filters, state variables, summations, derived calculations etc.), parameters (Monitoring and control thresholds, can be changed runtime) and events (Info, warnings and alarms) will be available as live data.

The turbine control system is running with a sample rate of 100Hz, but not all datapoints will be necessarily updated at every scan. Instead these are only updated if values change with respect to the previous scan or at least every 10 minutes.

3.6. Hybrid Twin

A Hybrid Twin will be developed for this test-bed using the complementary methodologies of Model Order Reduction (MOR) and system simulation. The former will make use of the results of high-fidelity (3D finite-element method, for example) simulations to generate a reduced order model that, once constructed, retains a high degree of accuracy, while itself requiring very little computational cost to execute. To consider the gaps between the model and the collected sensor data, referred to as "ignorance", ML (Machine Learning) techniques based on methods such as regression trees and decision trees will be applied on the fly.

The fault-modelling tools being developed as part of WP 3.3/3.4 will be exploited in this test-bed in order to develop a predictive maintenance application that will enable the wind farm operator to make informed decisions based on up-to-date estimates of the health of key WTG components.

Fault modelling analyses will be performed using methods based on both AI/MOR and system simulation. Reduced Order Models (ROMs) based on Proper Orthogonal Decomposition (PGD) and Topological Data Analysis (TDA) methods will be trained using historical operating data of the target WTG. They are well-suited to analysing high-frequency, oscillatory signals, such as the vibration data that will be collected from the sensors of the new yaw system. These ROMs will allow the construction of a database of common faults. Once the database is built, pattern recognition will be performed on the edge in order to identify the source of any possible fault in the operating turbine.

The system simulation approach to fault modelling commences with the construction of a system-level representation of the WTG based on nominal (non-faulty) operation. The system simulation paradigm allows for components of multiple physical domains (1D mechanics, 3D MBS [6]mechanics, hydraulics, thermal, electrical etc.) to be connected to construct a model that represents the whole physical asset and its constituent sub-systems in a physically realistic way. Once the nominal model is complete, functionality available in ESI's system simulation software, SimulationX, will be used to 'augment' it with components that model physical faults between components in the system, e.g. hydraulic leakages, electrical shorts, sensor fault, as well as parametric faults which model fault or degradation in the components themselves.

Depending on the amount of available historical sensor data for the target WTG, different ways of employing the fault-augmented model are anticipated. If little historical data is available, simulations of the model with different fault components (and their combinations) activated could be performed as a means of generating synthetic faulty data. This data could in turn be used to train ML models for detecting faults of the WTG in operation. If a larger quantity of historical data is available, in particular data that corresponds to real faults in the WTG, then the fault-augmented model could be dynamically calibrated using the incoming sensor data stream, thus providing a prediction of the 'amount' or locale of a fault in the real WTG.

A schematic representation of a nominal wind turbine model built using SimulationX is shown in Figure 9.

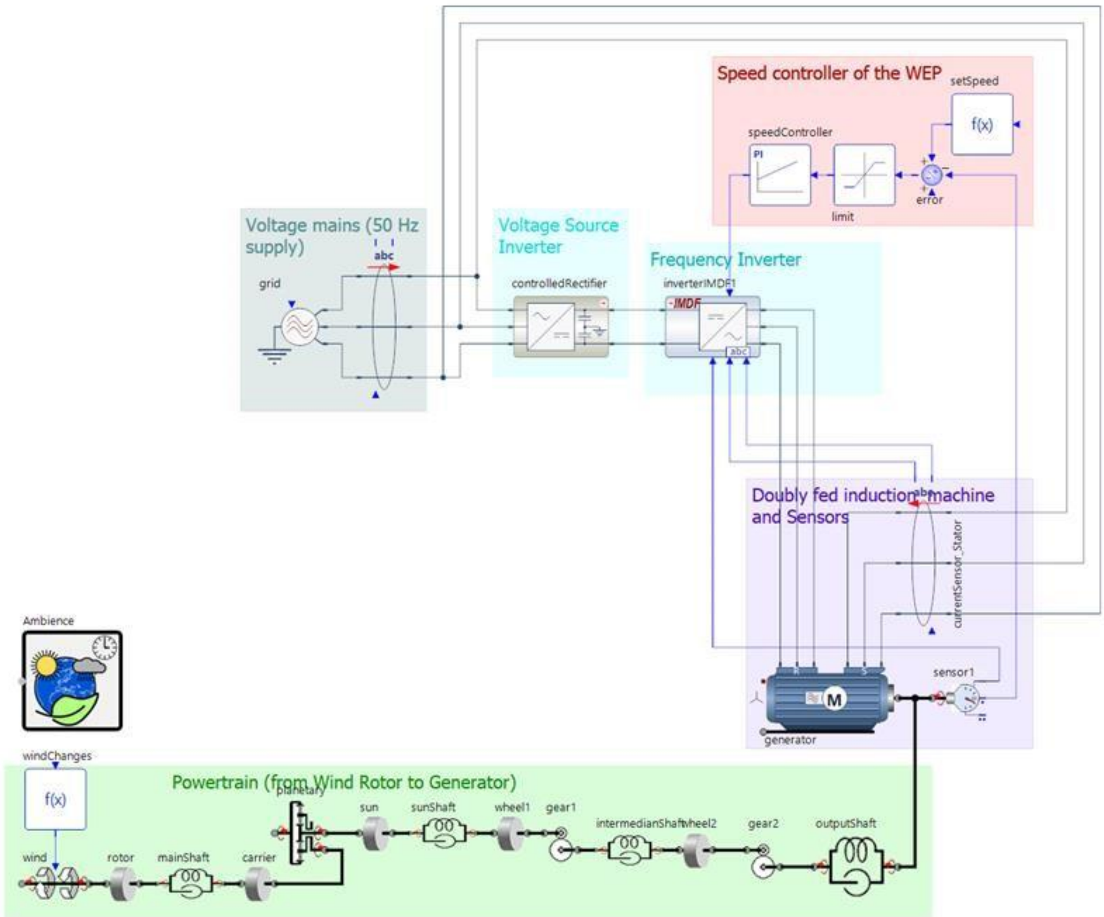


Figure 9: Representation of a nominal wind turbine model using SimulationX

Now that the target turbine has been identified, a system simulation model of the appropriate architecture will be constructed and parametrized using the technical characteristics listed in the previous section. Where

[6] Multibody simulation

parameters necessary to construct the model of the target WTG are not available, methods will be employed to calibrate the model based on the available operating sensor data.

In order to carry out the various fault analyses, simulations of key wind turbine sub-systems will be performed using both system simulation and 3D finite element simulation.

The 3D FEM simulations, will be performed on the cloud level in order to generate the results needed to construct the ROM. The resulting ROM will be executed on the edge, taking as input the necessary WTG sensor data to provide the operator with up-to-date estimates of the health of the system.

4. Test-bed 2: Machine tool spindle predictive behaviour, FILL

4.1. Identification of the use case

This test-bed is aimed at creating multiple target-oriented digital twins of machine tools (Syncromill) in a shop floor (esp. automotive series production). By deploying simulation and ML models of machine tools, drives and spindles for detecting the condition and behaviour of the spindle manufacturing-relevant and quality-influencing parameters (load, forces, vibrations etc.) are predicted. This enables the reduction of unexpected rejects, breakdowns and downtime by optimizing load and performance indices.

Involved stakeholders are the software architect for setting up the data collection, engineering experts for the set-up of experiments and data analysts for analyzing the data and developing models for prediction.

4.2. Technical preparation

Figure 10 shows the current architecture connecting the machine to an edge device. PROFINET[7] is an industry technical standard for data communication via Industrial Ethernet enabling data collection from controlling equipment in real-time (in the order of up to 1 ms). OPC Unified Architecture[8] (OPC UA) is a machine-to-machine communication protocol for industrial automation. PROFINET is suitable for controller-device communication, whereas OPC UA enables communication within controllers and builds the basis for data integration to the cloud. This architecture is currently available as well as the software architecture of *Cybernetics*. *Cybernetics, FILL's Industrial Internet of Things platform,* consists of components that are hosted on an TTTech-edge device in the control cabinet of FILL machines.

[7] https://www.profibus.com/technology/profinet/
[8] https://opcfoundation.org/about/opc-technologies/opc-ua/

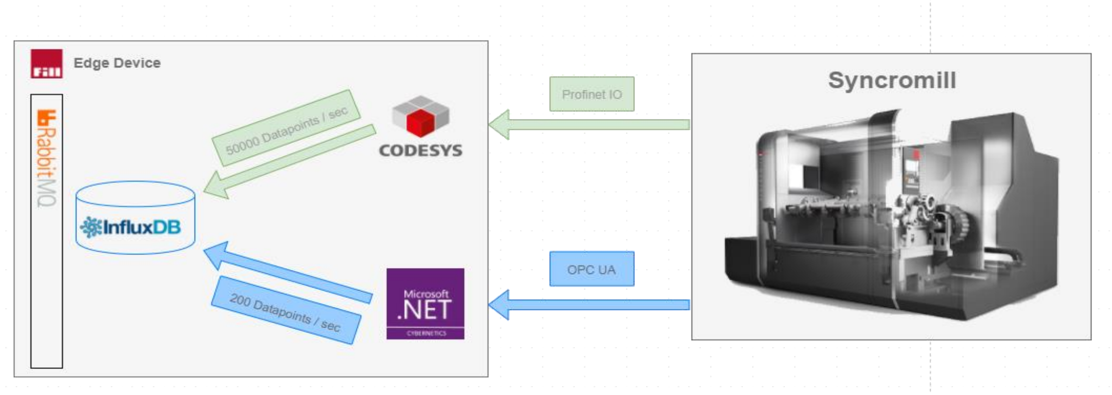


Figure 10: Data acquisition to the edge device

4.3. Infrastructure/Hardware to be used

Due to its competences FILL acts as an Engineering, Procurement, and Construction (EPC) contractor from sales, project management, mechanical engineering, safety equipment, software engineering, assembling to start of the operation. As a widely recognized supplier FILL implemented in the last years different machines up to large fully automated production lines for different industrial sectors.

In the following, details about the infrastructure to be used are given: Testbed 2 will be hosted in the FILL premises, which are shown in Figure 11.

Figure 11: FILL Future Zone – Area for Testbed 2

The standard machine tool (syncromill) for Testbed 2 is already set up and ready for implementation of IoTwins (see Figure 11). It is hosted within the Research and Development (R&D) area in the FILL Future Zone, as shown in Figure 13, and the edge device is built into the machine tool.

Figure 12: FILL Future Zone - Syncromill

4.4. Identification of sensors

To predict spindle and machine behaviour, a series of sensor data as well as calculated outputs from motor / logical controllers at different sample rates must be obtained.

A distinction is made between real time data ("fast channel") and non-real time data ("slow channel").

Table 6: Testbed variables extracted

Variable Name	Description	Sampling rate
Spindle 1 Nominal Position	Nominal Position for spindle 1 for each cycle calculated by the motor controller	166 Hz
Spindle 2 Nominal Position	Nominal Position for spindle 2 for each cycle calculated by the motor controller	166 Hz
Spindle 1 Actual Position	Actual Position for spindle 1 for each cycle measured by the motor controller	166 Hz
Spindle 2 Actual Position	Actual Position for spindle 2 for each cycle measured by the motor controller	166 Hz
Spindle 1 Torque	Current Torque for spindle 1 for each cycle measured by the motor controller	166 Hz

Variable Name	Description	Sampling rate
Spindle 2 Torque	Current Torque for spindle 2 for each cycle measured by the motor controller	166 Hz
Spindle 1 Power Consumption	Current power consumption for spindle 1 measured by the motor controller	166 Hz
Spindle 2 Power Consumption	Current power consumption for spindle 2 measured by the motor controller	166 Hz
Spindle 1 Velocity	Current velocity of spindle 1 measured by the motor controller	166 Hz
Spindle 2 Velocity	Current velocity of spindle 2 measured by the motor controller	166 Hz
Operating Mode	Current operation mode of the machine (automatic, manual, eg)	2 Hz

4.5. Data processing

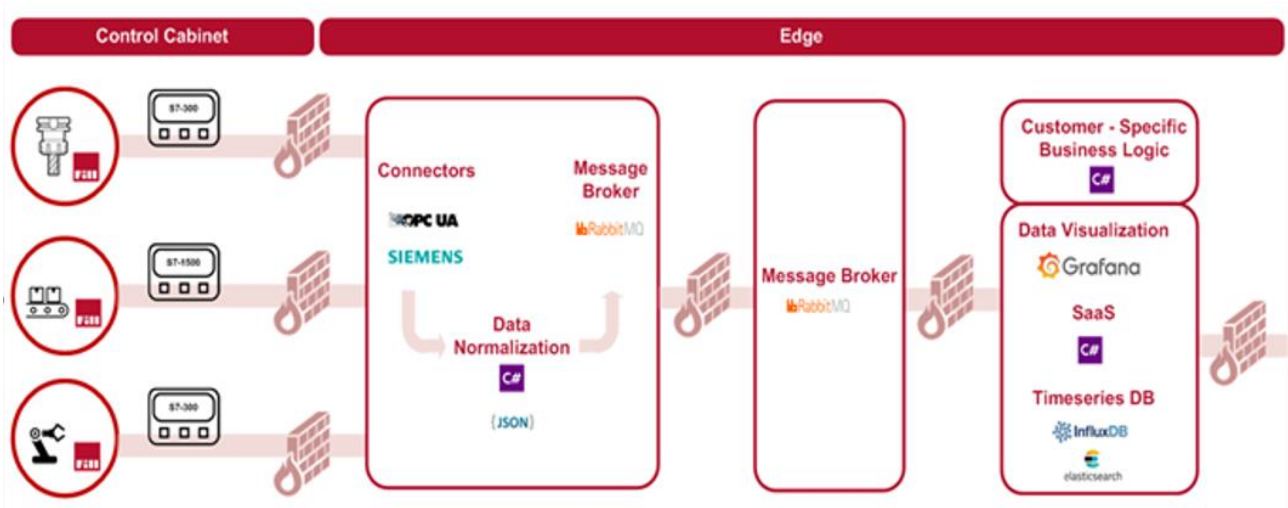


Figure 13: Components of Cybernetics for data processing

Figure 13 shows the individual components of *Cybernetics* and their relationships. The individual components of *Cybernetics* are:

- **Control Cabinet:** Automation is done using PLCs. *Cybernetics* connects to different PLCs and reads all relevant sensor, machine, and production data from these PLCs.
- **Connector:** *Cybernetics* implements connectors to PLCs. These connectors implement industry-standard protocols (e.g. OPC UA) but also proprietary protocols (e.g. the Siemens S7 protocol[9]).
- **Data Normalisation:** Every data point is normalised and converted to a standardised format, formatted as a JSON object.

[9] https://support.industry.siemens.com/cs/document/26483647/what-properties-advantages-and-special-features-does-the-s7-protocol-offer-?dti=0&lc=en-WW

- **Message Broker:** All data is distributed among all other components in *Cybernetics* using a message broker (RabbitMQ[10])
- **Customer-Specific Business Logic:** Varying customer requirements are implemented in customer-specific business logic modules.
- **Data Visualisation:** *Cybernetics* provides basic data visualisation and data analysis features on the edge device. Raw data is stored in a timeseries DB (InfluxDB[11]).

4.6. Digital twin

The aim of the digital twin is to supervise the spindle's behaviour. As soon as an unusual behaviour occurs, the machine should issue an alarm and report it. This involves the generation of an appropriate dataset, its analysis and the development of a machine learning model serving this purpose. Data from machine tools with additional sensory tool-holders is generated and milling tests for 3-4 weeks are processed. The evaluation of the experiments and data analysis as the starting point for machine learning models are currently work in process.

4.7. Validation/generalization

Validation is performed as a two-step process:

1. Milling-Experiments to generate data for proof-of-concept: Experiments with different parameters for data analysis and model generation for the digital twin. For this purpose, analyses are performed to investigate, whether the spindle behaviour can be predicted from spindle data.
2. Validation of the digital twin. Experiments are performed to induce anomalous spindle behaviour and to investigate and validate the performance of the digital twin model.

5. Test-bed 3: Crankshaft manufacturing system predictive maintenance, ETXE

5.1. Identification of the use case

High throughput crankshaft manufacturing system, the principal product of ETXETAR, is a semi-autonomous Computer Numerical Control (CNC) machine that produces an average of 1.000 crankshafts per day (a picture of such machine is provided in Figure 14). Downtime and breakdown reduction improve the productive efficiency of the whole line and consequence increase revenue. For instance, one of the main customers of ETXETAR estimates 50,000 €/hour for the cost of downtime in its production lines. Obviously, this figure would strongly depend on the process and on the type of company, but it seems necessary to work on it to contribute to the reduction of costs due to machine failures.

[10] https://www.rabbitmq.com/
[11] https://www.influxdata.com/

Figure 14: Etxetar's crankshaft manufacturing system

Within the IoTwins project, we studied the failure mode and effects analysis to identify the element with the highest Risk Priority Number (RPN). Having the highest RPN means that an unexpected failure in this element will stop the complete machine, affecting the complete production line. Therefore, the machine's element with the highest RPN is the frontal ball-bearing of the spindle head (see Figure 15). Frontal ball-bearing will be the focus for the testbed, producing high amount of data to generate a digital twin that will enable data-based maintenance, reducing the downtimes related to this part of the machine.

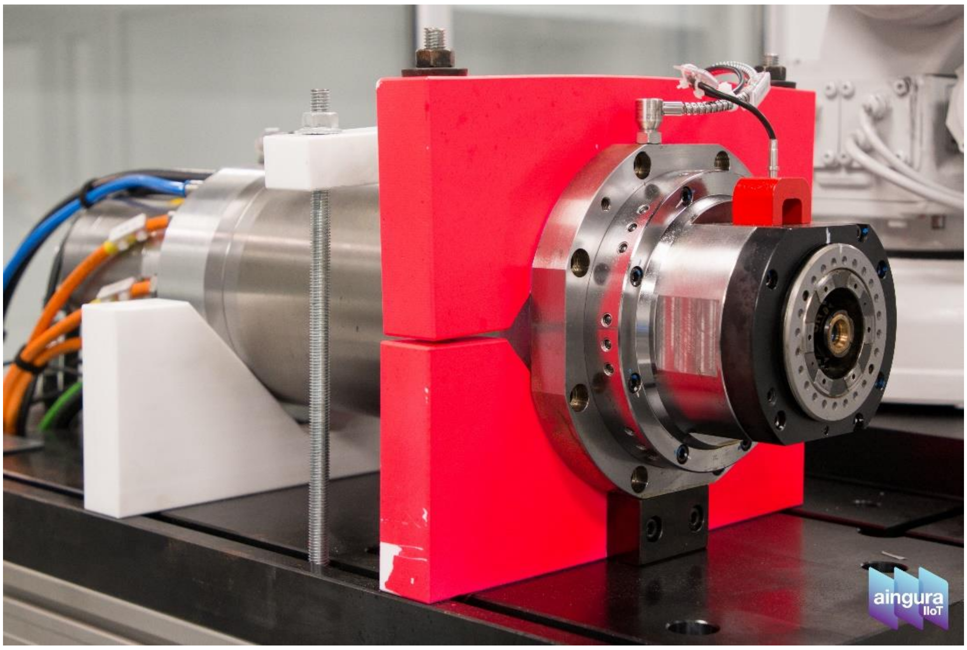


Figure 15: Crankshaft manufacturing spindle head (lab environment)

5.2. Technical preparation

This testbed is certified by Industrial Internet Consortium[12] (IIC). The Industrial Internet Consortium brings together the organizations and technologies necessary to accelerate the growth of the industrial internet by identifying, assembling, testing and promoting best practices. Members work collaboratively to speed the

[12] https://hub.iiconsortium.org/smart-factory-machine-learning

commercial use of advanced technologies. The testbed provides part of the infrastructure needed at three different deploying levels:

1. Lab: a synthetic setup with an infrastructure similar to the one deployed on the commercial products;
2. Factory-lab: a setup deployed on a real product when still in the factory (before deployment);
3. Production line: the actual final setup on the field and at scale.

These three scenarios will help to communicate, isolate (the ball-bearing analysis), test and validate all the required infrastructure related to data acquisition, pre-processing, and processing.

Therefore, at different deployment scenarios, the Aingura Insights (AI) edge computing node (see Figure 16), from our Linked Third Party, Aingura IIoT, S.L. will be used to extract different variables related to the frontal ball-bearing. Additionally, the AI will be used to run feature subset selection over those variables to send only the require variables, avoiding redundant and not related that will compromise posterior data analysis algorithms.

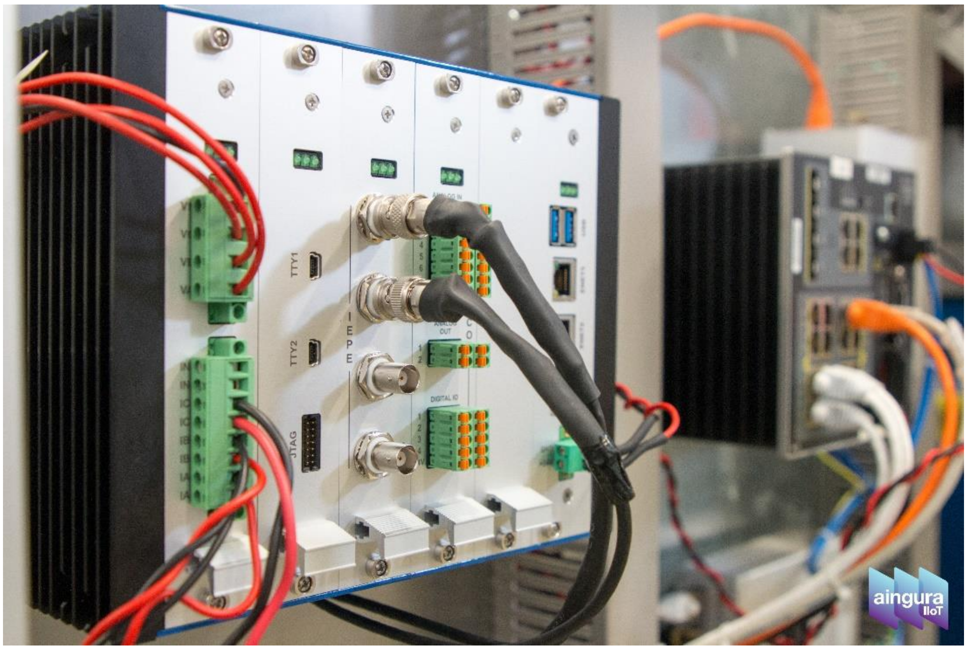


Figure 16: Aingura Insights (AI) edge computing node

5.3. Identification of sensors

The variables that will be extracted from different deployment scenarios are defined in Table 7.

Table 7: Testbed variables extracted and selected using AI

Variable Name	Description	Sampling rate
Raw accelerometer	Raw acceleration values coming from the signal acquisition	19.532 Hz
Ball-bearing fundamental train frequency (FTF)	AI's filtered acceleration amplitude value at FTF	0,33 Hz

Variable Name	Description	Sampling rate
Ball-bearing ball spin frequency (BSF)	AI's filtered acceleration amplitude value at BSF	0,33 Hz
Ball-bearing ball passing frequency at the inner cage (BPFI)	AI's filtered acceleration amplitude value at BPFI	0,33 Hz
Ball-bearing ball passing frequency at the outer cage (BPFO)	AI's filtered acceleration amplitude value at BPFO	0,33 Hz
FTF, BSF, BPFI, BPFO harmonics	AI's harmonics from the filtered signals of accelerometer amplitudes. The number of harmonics is open depending on feature subset selection results	0,33 Hz
Spindle head angular speed	Angular speed value extracted from machine's CNC	83,33 Hz
Spindle head torque	Torque value extracted from machine's CNC	83,33 Hz
Spindle head power	Power value extracted from machine's CNC	83,33 Hz
Spindle head current	Current value extracted from machine's CNC	83,33 Hz
Spindle head electrical windings temperature	Electrical windings temperature value extracted from machine's PLC	10 Hz
Spindle head cooling water temperature	Cooling water temperature value extracted from machine's PLC	10 Hz
Spindle head cooling water flowrate	Cooling water flowrate value extracted from machine's PLC	10 Hz
Spindle head frontal ball-bearing temperature	Frontal ball-bearing temperature extracted from machine's PLC	10 Hz

5.4. Data processing

There are two scenarios for data processing within the AI module:

1. Sanity check, filtering, imputation, sensor fusion. The aim of this step is to guarantee the data quality. The step will be performed in the AI module using its 4 CPU cores, but also, if acceleration is needed, its integrated FPGA. The most important task at this scenario is to extract the FTF, BSF, BPFI and BPFO as clean and fast as possible together with their harmonics. As the AI edge computing node does not have limits on the number of harmonics, a next step on variable selection must be used.
2. Feature Subset Selection (FSS). Variable selection is important to reduce the amount of data transmitted through a communication infrastructure. For example, accelerometer raw data can produce 800 MB/h, as it measures vibration from 3 to 10.000 Hz. However, depending on the analysis, only a few sets of frequencies are needed, drastically reducing those 800 MB/h to some kB/h. Furthermore, some of those frequencies could evolve when degradation occurs, being more

important than others, i.e., sometimes harmonics could provide more information than the fundamental frequency. This FSS could be applied using domain knowledge (expert on the field, for example, vibration engineer) and automatically, based on the evolution of data, that could be complementary to the field expert. The last case will be the approach in this testbed, testing online FSS algorithms running at the AI computing node.

5.5. Digital twin

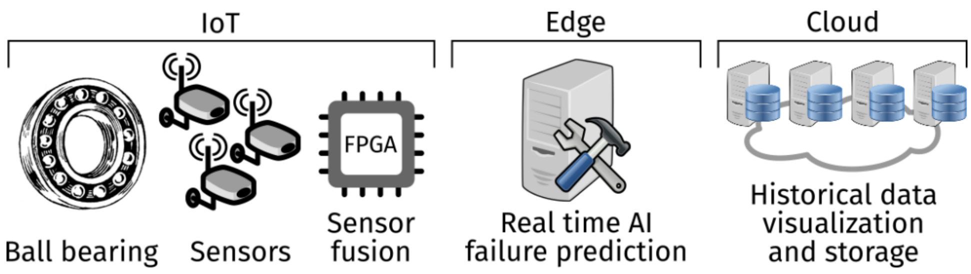


Figure 17: Dataflow of the testbed

The data collected by the sensors introduced in the previous sections are cleaned and fused by an FPGA-based IoT device (Aingura Insight). They are processed on the edge with AI techniques improving failure prediction. The testbed leverages the Cloud layer for storage and visualization of historical data. Figure 17 summarizes this dataflow.

5.6. Validation/generalization

As described before, the main aspect in this testbed is that is certified by the IIC with three different deployment scenarios. These three scenarios will help to create different validation steps, highly needed to isolate different characteristics of the solution before going into complete real world with several random events that could affect the system performance.

Therefore, the predictive maintenance system can be tested at lab phase to understand and validate if the algorithms and the hardware systems are doing what is expected, helping to debug implementation aspects. If an algorithm has low performance, it is better to adjust it in a controlled scenario, trying to reproduce different training examples that will help to generalize the system. Once the system is validated, the next step is to test under real machine scenario, with some added extra challenges, such as, unexpected turn on/off situations, different workpieces, etc. Here the complete system could be tested and adapted to have a clearer understanding of a real machine. The last phase will test the system within a real production environment, in an automotive OEM, having the influences and noise coming from the complete production line. In this case, the system should have the required performance independently of the scenarios that could create other machines in the line. For instance, if a machine is stopped because of a failure, the system must understand that it is a contextual issue that must be processed.

6. Test-bed 4: Predictive maintenance and production optimization for closure manufacturing, GCL

6.1. Identification of the use case

In the Spinetta Marengo (IT) Plant there are almost 200 machines with different purposes, brands, and type of use. With the support of the maintenance team and department managers, monitoring activities have been introduced to identify all possible unexpected events that should occur during the production life cycle and its possible related causes.

Among the various hypothetical failure events examined, it was decided to move towards a dedicated event of breakage which cyclically, but unpredictably, occurs on a defined plastic injection molding machine.

The anomaly that has been decided to prevent is an abnormal wear of the spindle bearing coating (worm screw that allows the mobile table to close by acting on the brace of the toggle). The material that comes off the bearing settles on the crests of the worm screw, damaging it.

6.2. Technical preparation

This machine identified was, at its state of the art, not connectable to any platform because of the lack of technical equipment (hardware and software). With the support of the machine supplier the machine has been equipped with the interfaces to expose externally its process data and the plant was predisposed with an appropriate networking structure in order to connect our plant machines with the edge server.

Moreover, PTC KEPServerEX [13]has been installed on the edge server, this software provides a series of drivers and plugin that allow the machine to communicate with our platform.

KEPServerEX is an industry connectivity platform that provides a single source of industrial automation data to all the applications. The platform design allows users to connect, manage, monitor and control different automation devices and software applications through one intuitive user interface.

Connected to KEPServerEX, an instance of another software, PTC ThingWorx[14], is responsible elaborating the normalized information returned, store and manipulate them into an historical database and send them to the webserver displaying human readable information to the end-user.

[13] https://www.kepware.com/en-us/products/kepserverex/
[14] https://www.ptc.com/en/products/iiot/thingworx-platform

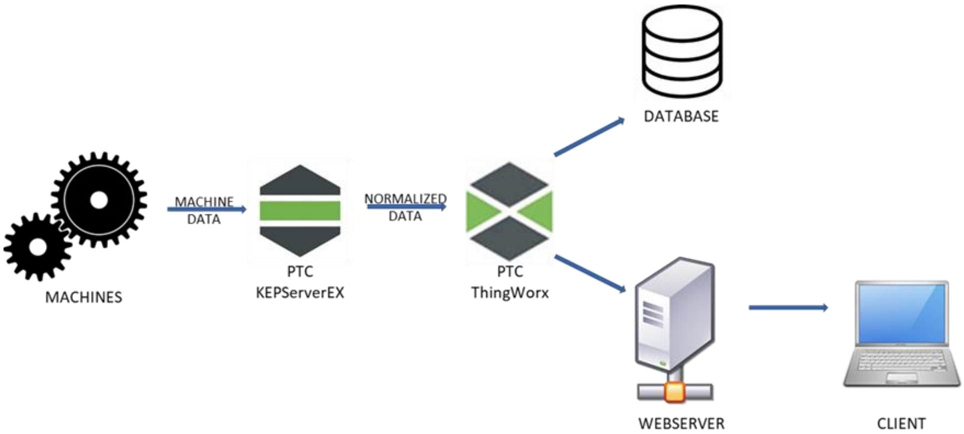


Figure 18: Data flow of the testbed

The molding machine chosen (see Figure 19) for the pilot natively provides about 2000 information tags, but just some of them are sensible information about the process:

Table 8: Identified important parameters from the molding machine

Description	Datatype	Scan Rate
Active Cycle	Long	500 ms
Active press	Long	500 ms
Automatic operating hours	Varchar	500 ms
Cavity Number	Long	500 ms
Cycle counter	Long	500 ms
Cycle Time Average	Float	500 ms
Cycle Time	Float	500 ms
Dosing time	Float	500 ms
Heating status	Long	500 ms
Heating zone cylinder 1	Float	500 ms
Heating zone cylinder 2	Float	500 ms
Heating zone cylinder 3	Float	500 ms
Heating zone cylinder 4	Float	500 ms
Heating zone cylinder 5	Float	500 ms
Heating zone cylinder 6	Float	500 ms
Heating zone cylinder 7	Float	500 ms
Heating zone cylinder 8	Float	500 ms
Injection Start	Float	500 ms
Injection Time	Float	500 ms
Material Cushion	Float	500 ms
Max pressure injection	Float	500 ms
Media housing temperature	Float	500 ms
Mold safety monitoring time	Float	500 ms
Mold safety time	Float	500 ms
Molded weight	Float	500 ms
Operating mode	Long	500 ms

Description	Datatype	Scan Rate
Press cycles	Varchar	500 ms
Press drive status	Long	500 ms
Press operating mode	Long	500 ms
Press status	Long	500 ms
Product Weight	Float	500 ms
Pump operating hours	Varchar	500 ms
Remaining material consumption	Float	500 ms
Scrap Pieces	Long	500 ms
Screw diameter	Float	500 ms
Semiautomatic operating hours	Varchar	500 ms
Switching pressure	Float	500 ms
Switching volume	Float	500 ms
Total weigh mould	Float	500 ms
Variable injection time	Float	500 ms

6.3. Identification of the sensors

To identify the correct breakdown event, it was necessary to find out the best way to obtain this information in real time. The solution adopted was to mount on the molding machine a series of new sensors to obtain real time information.

In particular:

- PLC Wago[15]: this tool allows the platform to retrieve information about the electric consumption, in particular:

Table 9: Parameters available from PLC Wago

Description	Datatype	Scan rate
Current Phase 1	Float	1000 ms
Current Phase 2	Float	1000 ms
Current Phase 3	Float	1000 ms
Total Active Power	Float	1000 ms
Total Apparent Energy	Float	1000 ms
Total Apparent Power	Float	1000 ms
Total Power Factor	Float	1000 ms
Total Reactive Power	Float	1000 ms
Voltage Phase 1	Float	1000 ms
Voltage Phase 2	Float	1000 ms
Voltage Phase 3	Float	1000 ms

[15] https://www.wago.com/us/discover-plcs

- Pruftechnik VIBGUARD COMPACT [16] Sensors: on Injection and Closing side these sensors have been placed to monitor rotation, temperature, and vibration of the bearings. For the injection bearing two warning thresholds has been set to detect the breakage event to prevent.

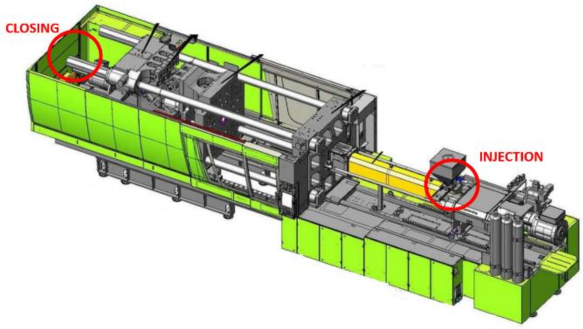


Figure 19: 3D-model of an injection moulding machine

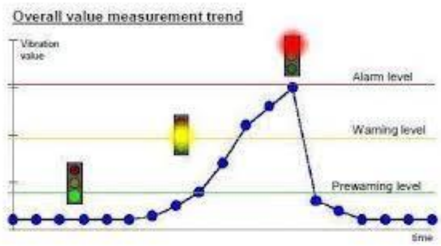


Figure 20: Overall value measurement trend

Table 10: Parameters of the VIBGUARD COMPACT sensors

Description	Datatype	Scan rate
Vibration Acceleration Closing Bearing	Float	1000 ms
Vibration Acceleration Injection Bearing	Float	1000 ms
Temperature Closing Bearing	Float	1000 ms
Temperature Injection Bearing	Float	1000 ms
Speed Closing Bearing	Float	1000 ms
Speed Injection Bearing	Float	1000 ms
Rotation Speed Closing Bearing	Float	1000 ms
Rotation Speed Injection Bearing	Float	1000 ms
Injection Bearing Status	Int	1000 ms

[16] https://www.pruftechnik.com/com/Products-and-Services/Condition-Monitoring-Systems/Online-Condition-Monitoring/Online-Condition-Monitoring-Systems/VIBGUARD-compact/

6.4. Data processing

At the scan rate configured on KEPServerEX all the information of the machine is saved into a SQL Database[17]. Every record is related to a specific registry and contains the timestamp of its saving. Every 30 minutes all the record contained in this database (more than 30K each run) are exported into a CSV file and stored locally on a server. At the same time, a subset of this information is continuously stored in a NoSQL database to allow the platform to process real time and batch statistics of the progress of some variables and to show the trends of the statistics taken into consideration in specific dashboards.

The information collected during the production cycle are available to the end-user through a dedicated user interface where both the progress of production and the monitoring of the process parameters are taken into consideration.

By the visual interpretation it will be possible to set which parameters to keep under control, define attention and alarm thresholds and correlate values/events that occurred during the production life cycle.

6.5. Digital twin

The aim for using a digital twin is to generate a production management optimisation framework. Here the expectation is to allow predictive maintenance based on big data combined with ML (machine learning[18]) algorithms to generate behavioural models[19]. On top of it, dynamic life monitoring algorithm will enable detect possible faulty conditions.

6.6. Validation/generalization

The validation aspect is done in one specific plant in Italy. To break down the overall goal to a limited number of machines in a controlled environment is reducing the complexity to understand the full behaviour of an injecting molding machine. The validation step is the comparison of the physical data to the behaviour model output and correct detection of a breakdown.

The aim for the generalization aspect is to port previous descripted approach to other machines in the plant/to other plants in the world. Therefore, the following points are important: connectivity, data access, scalability, company-wide device-to-cloud connectivity.

[17] https://embnet.vital-it.ch/CoursEMBnet/Basel07_II/Introduction%20to%20Database%20Systems,%20Data%20Modeling%20and%20SQL.pdf

[18] Ayodele, Taiwo. (2010). Machine Learning Overview. 10.5772/9374.

[19] Rahwan, Iyad & Cebrian, Manuel & Obradovich, Nick & Bongard, Josh & Bonnefon, Jean-François & Breazeal, Cynthia & Crandall, Jacob & Christakis, Nicholas & Couzin, Iain & Jackson, Matthew & Jennings, Nicholas & Kamar, Ece & Kloumann, Isabel & Larochelle, Hugo & Lazer, David & McElreath, Richard & Mislove, Alan & Parkes, David & Pentland, Alex & Wellman, Michael. (2019). Machine behaviour. Nature. 568. 477-486. 10.1038/s41586-019-1138-y.

7. Conclusion

This deliverable reports the outcome of the work done within subtask 4.1.1, 4.2.1, 4.3.1 and 4.4.1 to smartify or sensorize the systems related to the testbeds and data acquisition. In addition D4.1 concentrate on enhanced data collection and integration for the manufacturing testbeds. As one conclusion cannot be generalized for the deliverable, it is sectioned by testbeds.

Testbed 1 (BRI) focuses on deploying the sensorized WTG to the end-user in order to realize the high-volume data transmission from YAW and SCADA system. BRI is setting a new architecture in order to merge to different monitoring systems. By defining Cloud-2-Cloud architecture, it allows the partners receiving and using data for the TB1 purpose. Based on historical datasets for starting a preliminary modelling approach, which is propaedeutic to the hybrid Model, the testbed will run this model on the edge computer. Ecosystem data will be considered for the improvement of the model and the optimization of the WTG energy production.

In the future, investigation of a new control strategy of the entire wind park aims to reduce maintenance cost and optimize WTGs usage.

Testbed 2 (FILL) concentrates on modelling predictive spindle behaviour of the standard machine tool *SYNCROMILL*. Therefore, redundant IoT modules for logging data in high sample rate are designed and installed. Sensors are identified and data is processed in two frequencies: 2Hz and 166Hz. Experiments with additional sensory toolholders for data generation are performed. Currently evaluation and data analysis are under development.

Future work contains data analysis and validation and whether and how spindle behaviour can be predicted from spindle data.

Testbed 3 (ETXE) in collaboration with BSC is focusing on enabling data-based maintenance, reducing the downtimes related to the ball-bearing of semi-autonomous crankshaft manufacturing system produced by ETXE-TAR. The testbed is certified by Industrial Internet Consortium (IIC) and the infrastructure is organized on three levels: laboratory, factory-laboratory, and production line. The first phase consists in laboratory deployment. The following achievement were recorded:

- Identified 13+ variables sampled at 4 different frequencies corresponding to data gathered from sensors deployed on the laboratory setup.
- Basic sanity check, filtering, imputation, sensor fusion has been implemented on the IoT side.

Future work consists of improving acquisition, feature subset selection is on-going on a multicore embedded computational unit deployed on the edge and integration of the cloud infrastructure.

Tested 4 (GCL) focuses on modelling predictive maintenance behaviour of a particular component related to molding machines. In order to forward sensor data to a remote platform, new modules sensors were installed in the machine to acquire and logging sensible data remotely at two timesteps: 500ms and 1000ms. On edge level, an industry connectivity platform has been installed to retrieve and normalize information, stored in a local SQL database and exported in files, from machine modules.

Future work aims implementing the digital twin platform in order to visualize machine information and better identify and model standard and not-standard behaviours.

8. Comments

As the outbreak of virus COVID-19 (also known as SARS-CoV-2[20]) as a global health crisis is touching all segments of social, economic and human aspects, the pandemic is also affecting the progress of the IoTwins project. Due to the ongoing situation, different countries have put/are putting different restrictions to the people living in the respective countries or traveling to it. Even regions or countries were put/are in lockdown to protect the people.

WP4 consists of four large-scale industrial testbeds in the manufacturing sector. Consequently, it requires the purchase and delivery of hardware, the setup/installation/ commissioning of the smart systems, training of the personnel, testing of the IoTwins platform and analysing of the data.

The following activities related to this deliverable were affected by the pandemic:

- Delays in the production and/or delivery by the suppliers for hardware (e.g. sensory tool-holders)
- Manufacturing of required pieces in the plants
- Reduced possibilities to have workshops and therefore optimized timing
- Meetings with end-user/clients
- The work package related to the digital twin infrastructure was delayed resulting in a not testable system at this point
- The training of the personnel to install the sensorized system takes more time due to the social distancing

On the other side, communication (digital) between persons, evaluation of the result and analysis is not affected by the pandemic. Early work on data acquisition in the laboratory, like in the case of ETXE and BSC, has not been affected by the COVID-19 outbreak. This allowed to have enough data to continue working isolated from the rest of the project on variable identification and sensor fusion during the lockdown.

[20] https://www.who.int/health-topics/coronavirus