MANUFACTURING DATA NETWORKS

Title	Industrial pilot area validation & pilot benchmark and KPIs_Process Operations
Document Owners	POLIMI
Contributors	ENG, CERTH, Atlantis, AVIO, CNR, +GF+, FRAISA, Siemens, DataPixel, SSF, Core, MIR
Dissemination	PUBLIC
Date	30.09.2025
Version	V1

Table of contents

List of Figures

List of Tables

Document Status

Deliverable Leader	POLIMI
Internal Reviewer 1	UNINOVA
Internal Reviewer 2	NOVA
Work Package	WP5
Deliverable	D5.3: Industrial pilot area validation & pilot benchmark and KPIs_Process Operations
Due Date	M40
Delivery Date	30.09.2025
Version	V01

Version History

0.1	Nima Rahmani Choubeh (POLIMI); Table of Contents
0.2	Pietro Greco (ENG); Section 3
0.3	Roberto Perez (+GF+); Section 2
0.4	Nima Rahmani Choubeh (POLIMI); Section 4
0.5	Nima Rahmani Choubeh (POLIMI); Section 1, Section 5
0.6	Pietro Greco (ENG), Roberto Perez (+GF+); Section 4 (validation)
0.7	Erica Perego (MIR); Section 2, Section 3 (validation)
0.8	Nima Rahmani Choubeh, Walter Quadrini (POLIMI); Proof-check reading, content harmonization
0.9	Ruben Costa (UNINOA): First Internal Review
1.0	NOVA: Second Internal Review
2.0	Nima Rahmani Choubeh, Walter Quadrini (POLIMI); addressing reviewers' comments, finalisation

Further Information

More information about the project can be found on project website: https://re4dy.eu/

Disclaimer

Project Partners

Number	Participant organisation name	Acronym
1	ASOCIACIÓN DE EMPRESAS TECNOLÓGICAS INNOVALIA	INNO
1.1	CBT COMUNICACION & MULTIMEDIA SL	CBT
2	CHALMERS TEKNISKA HOGSKOLA AB	Chalmers
3	INTERNATIONAL DATA SPACES EV	IDSA
4	VOLKSWAGEN AUTOEUROPA, LDA	VWAE
5	ASSECO CEIT AS	CEIT
6	UNINOVA-INSTITUTO DE DESENVOLVIMENTO DE NOVAS TECNOLOGIAS-ASSOSIACAO	UNI
7	FILL GESELLSCHAFT MBH	FILL
8	AVL LIST GMBH	AVL
9	VISUAL COMPONENTS OY	VIS
10	UNIVERSIDAD MIGUEL HERNANDEZ DE ELCHE	UMH
11	ATLANTIS ENGINEERING AE	ATLANTIS
12	DATAPIXEL SL	DATA
13	CORE KENTRO KAINOTOMIAS AMKE	CORE
14	UNIVERSITETE I OSLO	UiO
15	GE AVIO	AVIO
16	ENGINEERING-INGENIERIA INFORMATICA SPA	ENG
17	POLITECNICO DI MILANO	POLIMI
18	ATOS IT SOLUTIONS AND SERVICES IBERIA SL	ATOS IT
18.1	ATOS SPAIN SA	ATOS ES
19	KATHOLIEKE UNIVERSITEIT LEUVEN	KU
20	NETCOMPANY-INTRASOFT SA	INTRA
21	NOVA ID FCT - ASSOCIACAO PARA A INOVACAO E DESENVOLVIMENTO DA FCT	NOVA
22	INDUSTRY COMMONS FOUNDATION (INSAMLINGSSTIFTELSE)	ICF
23	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	CERTH
24	GRUPO S 21SEC GESTION SA	S21SEC
25	UNIVERSITAT POLITECNICA DE VALENCIA	UPV
26	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR
27	SOCIEDAD ANDALUZA PARA EL DESARROLLO DE LAS TELECOMUNICACIONES SA	SANDETEL
28	SWITZERLAND INNOVATION PARK BIEL/BIENNE AG	SSF
29	GF MACHINING SOLUTIONS AG	GFMS ADVMAN
30	FRAISA SA	Fraisa SA
31	SIEMENS SCHWEIZ AG	SIE
32	MIRAITEK SRL	MIR

Executive Summary

This document reports the comprehensive industrial pilot area validation, benchmarking, and KPI assessment related to process operations within the RE4DY project. It focuses on full-scale implementations, industrial trials, and performance monitoring of two major pilots GF Fraisa and Avio Aero. Each pilot integrates advanced AI and digital technologies following the RE4DY reference architecture to address key business scenarios such as tool selection and virtual process preparation, tool lifetime prediction through federated learning, machine maintenance via predictive analytics, and adaptive digital manufacturing using in-process metrology.

GF Fraisa pilot demonstrates successful integration of the FRAISA ToolExpert with Siemens NX CAM for tool and process preparation, yielding reduced setup times, fewer errors, and optimized energy consumption. Advanced AI applications developed by CORE and Atlantis leverage on federated learning to deliver predictive maintenance and tool wear prediction, achieving up to 80-83% accuracy and promising further improvements. Machine maintenance applications monitor critical components with AI models to predict failures, enhancing machine uptime by 10-15% and reducing maintenance costs by up to 30%. The Adaptive Digital Manufacturing pilot employs in-process metrology, enabling closed-loop control that slashes machine verification time and production scrap rate, while improving cycle times.

Avio Aero pilot introduced automated defect detection leveraging state-of-the-art deep learning (YOLOv8) under constrained datasets. Despite data limitations and annotation challenges, models exhibit potential for effectively detecting even subtle surface defects. The pilot also includes explainable AI features and a cognitive training suite to enhance inspector performance and reduce training hours. Furthermore, predictive quality and maintenance frameworks based on federated learning have been deployed successfully, delivering measurable impacts on Overall Equipment Effectiveness (OEE) in manufacturing using EDM machines.

The 6P Performance Monitoring Framework is applied to systematically monitor digital maturity and performance impact across technical and socio-business dimensions in both pilots. Survey and interview results reveal significant progress in process integration, data sharing, federated learning adoption, and KPI achievement, although barriers such as cybersecurity constraints, data standardization, and infrastructure heterogeneity remain. Lessons learned emphasize the importance of standardized data models, robust digital threads, scalable federated learning architectures, and sustained cross-organizational collaboration.

1 Introduction

Context and scope of this document

This document reports the industrial pilot area validation of work package five of RE4DY project. The full scale-up implementation of the industrial pilots (AVIO AERO and GF) has been fully described in this deliverable including establishment of final architecture and integration of it to the industrial environment as well as reporting the revised KPIs related to each business scenario leveraging on RE4DY reference architecture. In section 4 the outputs of task 5.4 has been depicted introducing the POLIMI performance monitoring methodology and its two iterations and insights of the pilots on project concepts and reference architecture.

Relationships among other deliverables

This deliverable is closely related to D5.2 "Scale up & on-site validation & revised KPI assessment: Process Operations" and its related deliverables in WP2 and WP3 of the project. In addition, this document is well connected with D4.3 of WP4 titled "Industrial pilot area validation & pilot benchmark and KPIs_process engineering" and some are of the section are closely aligned.

2 Pilot 3: GF Fraisa

General Introduction

The GF Fraisa pilot implements services for virtual machining preparation, tool lifetime and machine maintenance management as well as part quality control and overall optimisation for the case of milling technologies. The pilot is centred on the machine and tools, but as the scenario is deployed across the tool and machine lifecycle for high productivity and high precision applications, business process related to virtual planning and adaptive manufacturing and quality control are included. The challenges addressed are the following:

1. Selection of best tools for a given part manufacturing, with virtual simulation of manufacturing KPIs
2. Individual tool lifecycle management with AI prediction of tool wear for optimized tool recycling
3. Predictive maintenance of key machine components for guaranteeing high precision and maximize uptimes
4. On machine quality control of manufactured parts for adaptive manufacturing

Those challenges are associated with the corresponding business processes (BP):

BP1 – Process Planning and Preparation

- Objective: Tool information available with CAM and machine conditions for process planning & simulation.
- Benefit: Selection of best tools and strategies for optimized machining processes.

BP2 – Tool Management and recycling

- Objective: Tool data integration for machine operation and Monitoring of tool status and timed recovery and refurbishing of tools with predictive solutions.
- Benefit: increased recycling and timely refurbishing of tools via predictive insights

BP3 – Machine Maintenance

- Objective: Maintenance of critical machine components.
- Benefit: Monitoring of component status and timed warning, repairing or refurbishing process with predictive solutions.

BP4 – Adaptive Digital Manufacturing

- Objective: Machine Verification using metrology and advanced part alignment.
- Benefit: Automated in-machine metrology and feedback.

The following picture (Figure 1) represents the architecture of the pilot for all the business processes, requiring specific modules related to federated learning FEDMA and FPdM as specific application for predictive maintenance.

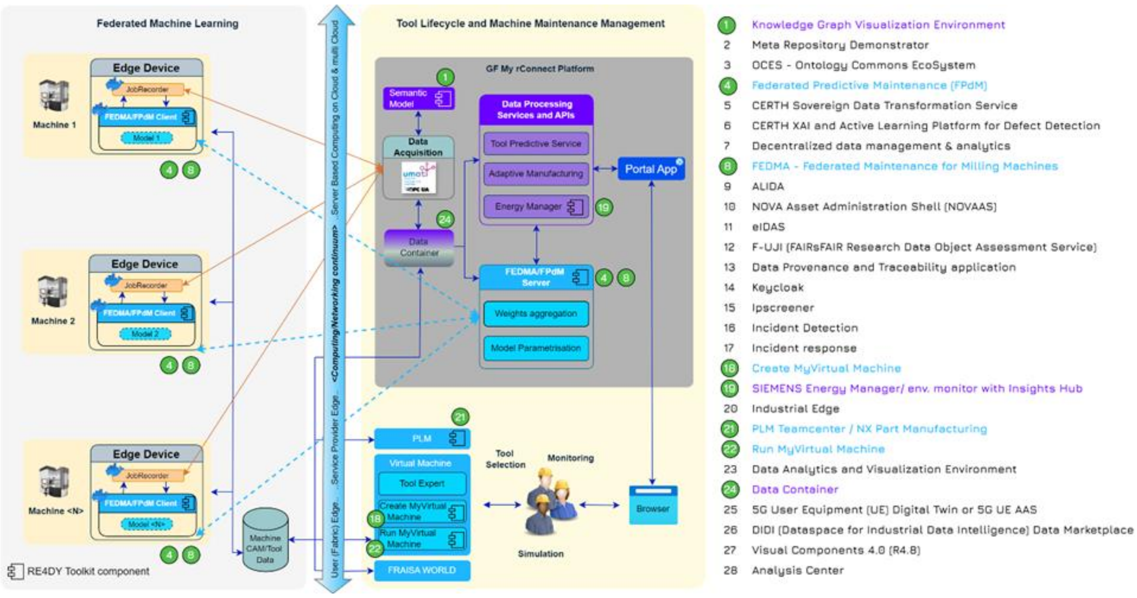


Figure 1 – General architecture of GF pilot showing the RE4DY toolkit elements deployed across four business processes

The Data Container

A central achievement of the pilot was its ability to bridge and connect previously isolated data silos, solutions, and processes across the machine tool ecosystem. The core motivation behind this effort is simple: by bringing data from multiple sources into a unified environment, stakeholders can unlock new types of value-added services that would be impossible to implement in isolation. For example, the work done by Atlantis and Core demonstrates how combining machining and tool data enables advanced predictions about tool and machine condition.

Rather than functioning as a traditional data container in the strict sense, the solution developed acts more as a data aggregation and orchestration layer. Its design is flexible, data storage and access can be adapted based on the specific requirements of a use case, whether that involves a data connector, a container, or a marketplace interface.

In the current setup, its primary role is to interface with various data sources: reaching edge devices on machines to extract operational data, identifying which tools were used via integrated solutions, retrieving corresponding part file information, and finally aggregating metrology and quality control data at the end of the process. This comprehensive aggregation enables researchers to conduct end-to-end analysis: for example, examining how specific process conditions influence tool wear, and how that wear, in turn, affects final part tolerances.

This holistic view, from engineering through to quality control, is the key value of the system, enabling full traceability and insight across the production process. Looking ahead, it would be valuable to expand this scope further by incorporating lifecycle data of the product post-factory or during its return, closing the loop for circular manufacturing.

2.a Business Scenario 1:

This business scenario focuses on the virtual preparation of the machining process, integrating the Fraisa ToolExpert application in the NX CAD CAM for the best choice of the tool configuration given a 3D model specification, and the virtual manufacturing of the part, which helps to avoid collisions and verify if the full program is consistent with the requirements prior to machining.

2.a.1 Full-scale implementation

ToolExpert integrated in Siemens CAM NX

FRAISA ToolExpert is now seamlessly integrated into Siemens CAM NX as shown in Figure 2. Initially, users could transfer tool geometry data with just a few clicks – now, cutting data can be transferred just as easily. The FRAISA ToolExpert is now an integrated vendor in the Cloud Connect Tool Manager in Siemens NX.

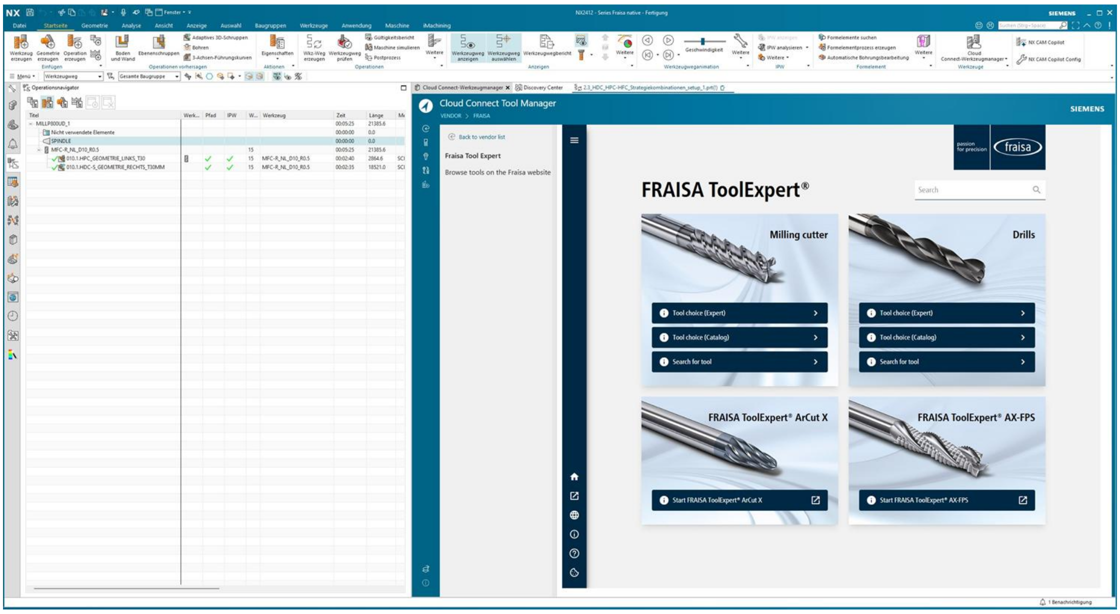


Figure 2 – ToolExpert integrated in Siemens NX

Once the tool is chosen the geometrical tool data can be sent to the CAM with one button as can be seen in Figure 3.

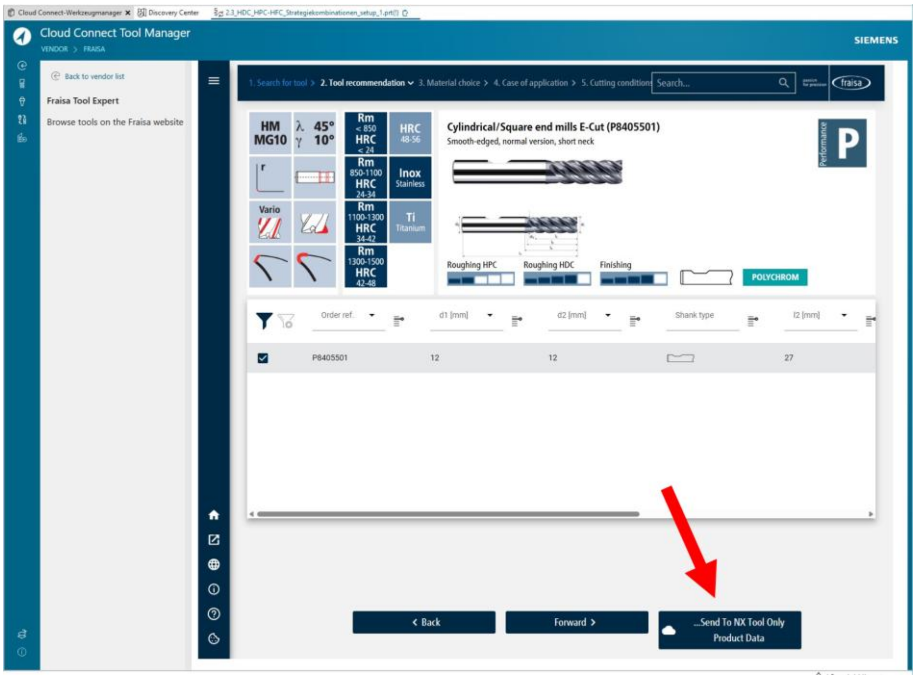


Figure 3 – Geometrical data send to NX Tool Only Product Data

If the workpiece material and milling strategy are chosen as well, the recommended cutting data can be sent to the CAM as well (Figure 4).

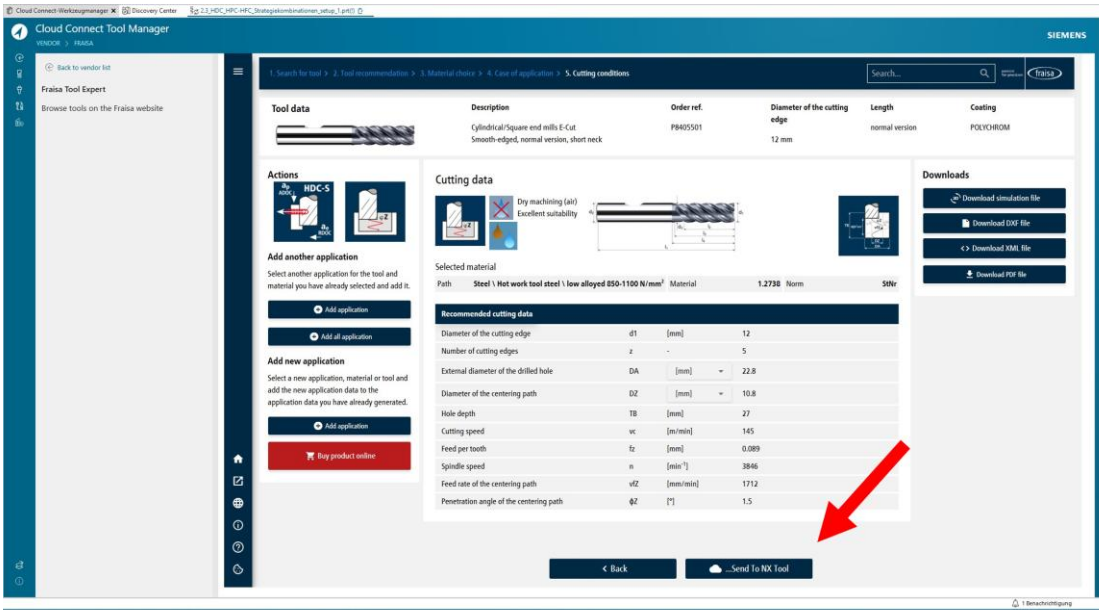


Figure 4 – Cutting data Send to NX Tool

When the tool is sent to the CAM tool-library, the package (tool geometry data and cutting data) can be transferred to the NX jobs. The tool in the tool-library is shown in Figure 5.

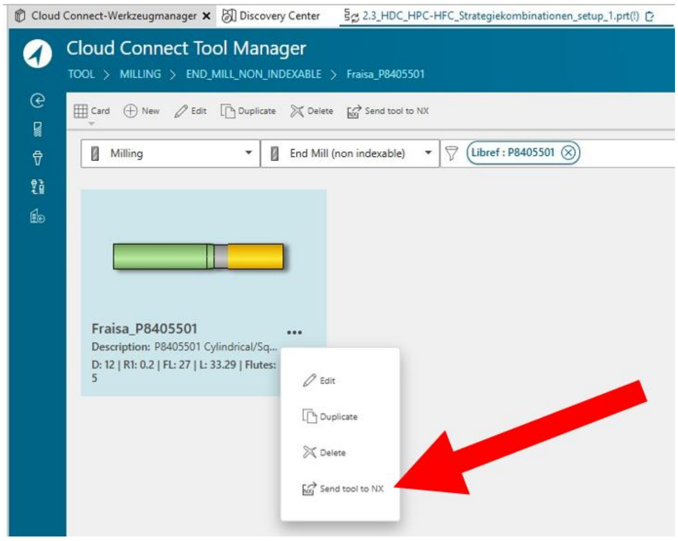


Figure 5 – Transfer tool geometry data and cutting data to NX jobs

Afterwards, the tool with the exact geometrical data and the recommended cutting data is in the Operations Navigator from NX (Figure 6).

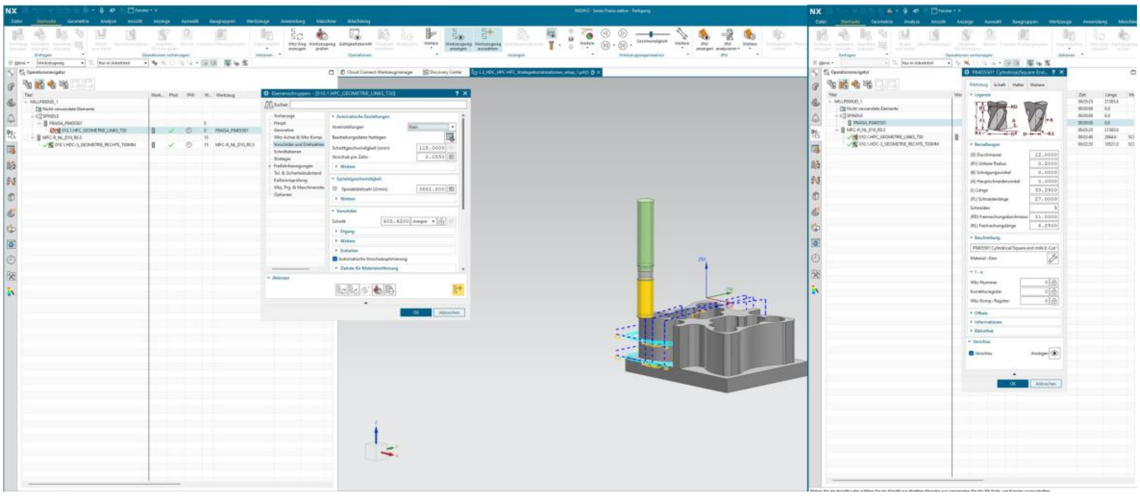


Figure 6 – Geometrical data and cutting data in Operations Navigator

In the example of the RE4DY drone component, four tools with geometry and cutting data must be integrated into the CAM system. Thanks to the integrated ToolExpert, this can be done quickly and reliably.

Virtual Environment

As illustrated in Figure 7, virtual simulation, conducted within the Run MyVirtual Machine or Create MyVirtual Machine software environments, integrates the designed workpiece geometry, the specified cutting tool from Fraisa's ToolExpert, and the CAM-generated NC program. This simulation provides crucial insights into the projected machining duration, considering the chosen tooling and the defined preparation strategy. Furthermore, it enables the identification of potential kinematic interferences or collisions between the tool, the workpiece, and the machine tool housing, thereby enhancing process planning and mitigating risks

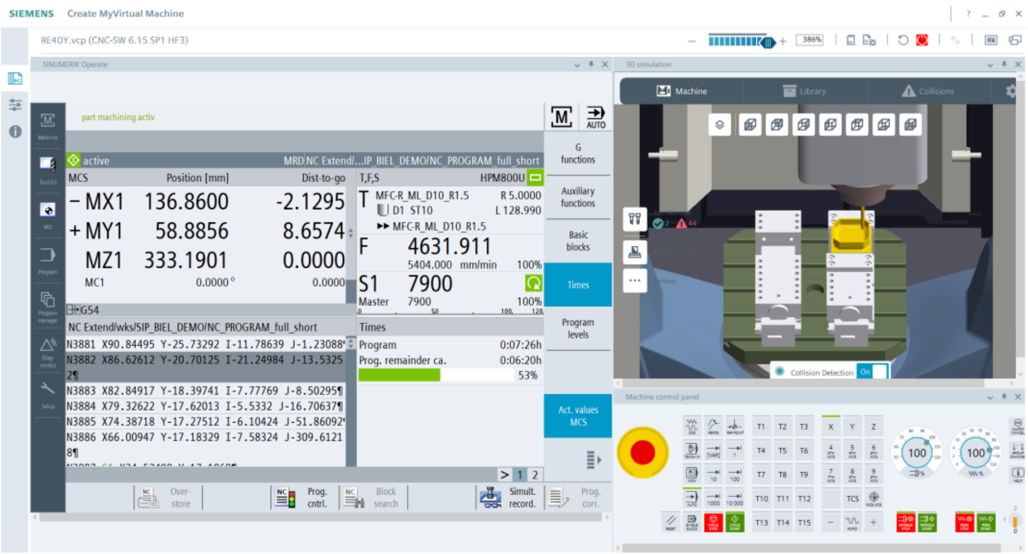


Figure 7 – Milling simulation in Create MyVirtual Machine

Moreover, Figure 8 demonstrates that the energy consumption required for processing work piece can be assessed within the simulated manufacturing environments of Create MyVirtual Machine or Run MyVirtual Machine.

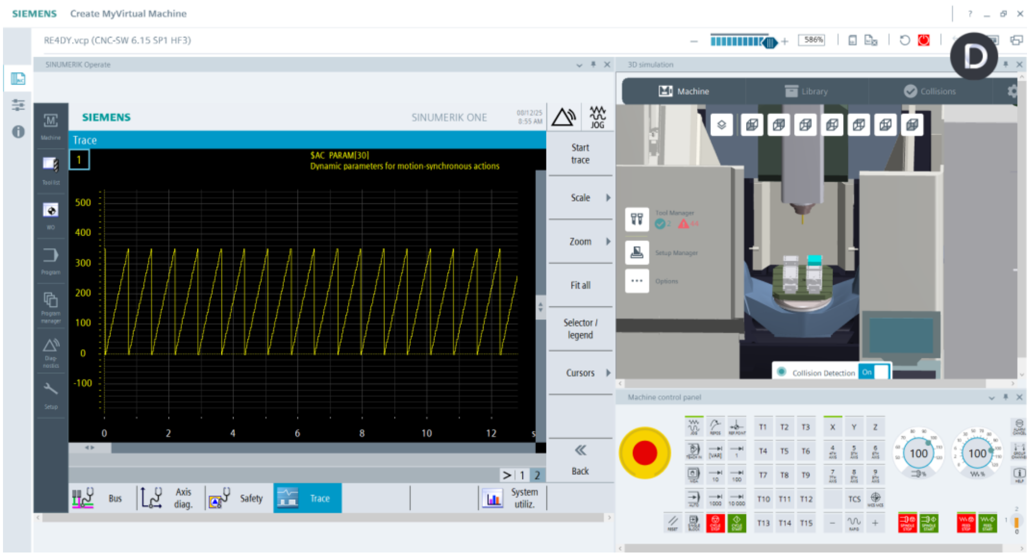


Figure 8 – Trace of energy consumption in Create MyVirtual Machine

2.a.1.1 Architecture

Figure 9 illustrates the architecture for the virtual commissioning of a part. This framework integrates the Fraisa ToolExpert within the CAD/CAM software environment, enabling precise tool selection for manufacturing processes. The defined machining operations are subsequently simulated using Siemens Run MyVirtual Machine.

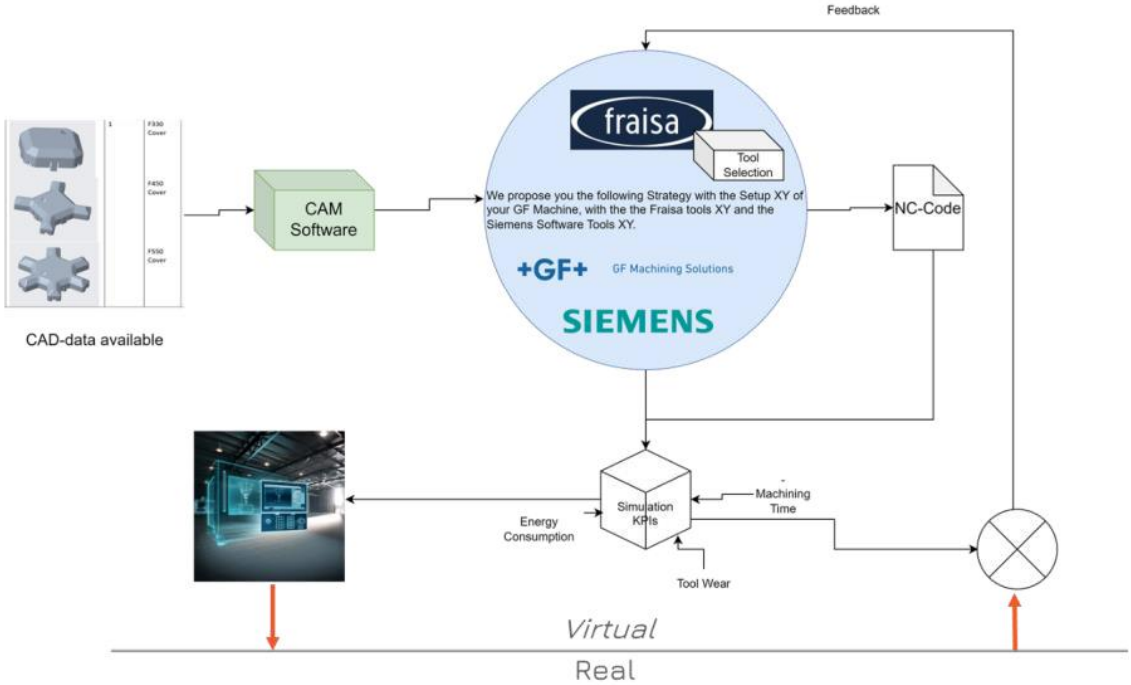


Figure 9 – Architecture for Virtual commissioning

2.a.1.2 Applications

ToolExpert

The FRAISA ToolExpert is a digital tool designed to quickly and accurately determine cutting data for FRAISA tools. It supports various machining processes such as milling, drilling, and trochoidal machining by calculating optimal parameters like cutting speed, feed rate, step-over, and spindle speed based on the selected material, tool type, machining strategy, and machine setup.

With an intuitive user interface and a comprehensive database of materials and tools, the ToolExpert helps improve efficiency, reduce tool wear, and optimize process times. The generated data can be seamlessly integrated into CAM systems like Siemens NX, making it an ideal solution for modern, high-performance manufacturing environments.

Run MyVirtual Machine

Siemens Run MyVirtual Machine is a comprehensive software solution designed to accurately simulate the entire machining process on a virtual representation of a CNC machine. It supports the validation of Numerical Control (NC) programs and machine kinematics by calculating critical performance indicators such as estimated machining time, energy consumption, and potential tool wear, based on the loaded NC code, machine configuration, and part geometry.

With its ability to precisely replicate real-world machine behavior, Run MyVirtual Machine helps identify and prevent potential collisions, programming errors, and inefficient movements before any physical material is cut. This robust virtual environment allows for the thorough evaluation and optimization of machining strategies, significantly improving efficiency, reducing costly physical prototypes, and minimizing machine downtime. This virtual validation process benefits from precise tool and cutting data selection, such as those provided by Fraisa's ToolExpert within CAM systems like Siemens NX, creating a holistic digital workflow from design to validated production.

2.a.1.3 Key challenges and solutions for full-scale implementation

In the future, the customer should no longer have to manually specify the material, as it should already be defined upon importing the CAD file. The milling strategy, along with suitable cutting parameters, should also be automatically suggested based on the part or the specific area to be machined. To enable this, the CAD model must be equipped with the corresponding macros.

The milling strategy should be optimally determined based on conditions such as the part material, pocket depth, pocket size, pocket corners, the milling machine, and the available tools (whether new or already showing wear). Currently, CAD features are either not recognized at all or only partially. Tool suggestions are made solely based on the material to be machined, while the strategy still has to be selected manually.

To automate this process, algorithms are missing that can determine the appropriate strategy and parameters based on the above-mentioned features.

2.a.2 Industrial trials of the pilot

2.a.2.1 Testing procedure and Barriers

This section describes the testing procedure and technical activities undertaken to implement the proposed use case, focusing on the integration of Fraisa's ToolExpert with Siemens NX and Siemens' virtual manufacturing environments. Furthermore, it addresses the specific barriers encountered during these tests on industrial equipment and pilot setups, along with the measures adopted to mitigate them or improve the expected output.

- CAD Model Creation: The initial phase involved the design of the drone cover within Siemens NX. This CAD model served as the foundational element for subsequent manufacturing simulations. Detailed geometric modeling was performed to accurately represent the workpiece.
- Virtual Environment Configuration: A critical step was the configuration of the virtual manufacturing environment. This necessitated the creation of a digital twin of the target machine tool, specifically the GF MillP800US, within Create MyVirtual

Machine. This digital representation served as the platform for integrating the generated NC code and the tool definitions.

- Tool Data Integration: Tools selected for the machining process were sourced from Fraisa's ToolExpert database. These tool definitions, including their geometries and cutting parameters, were subsequently integrated into the virtual environment.
- NC Code Generation and Simulation: Following the tool integration, NC code was generated based on the designed workpiece and selected tools. Simulations were then conducted with this NC code in Create MyVirtual Machine (for setup and programming verification) and Run MyVirtual Machine (for real-time simulation and optimization). This facilitated the validation of the machining process, assessment of machining duration, and identification of potential collisions. During the implementation and testing phases, several technical and logistical barriers were encountered.
- Software Licensing Barrier: Key software components, including Siemens NX, Create MyVirtual Machine, and Run MyVirtual Machine, are proprietary, license-based solutions that require on-premises installation. This necessitated significant lead time for procurement, license management, and IT infrastructure setup.
- Digital Twin Creation and Validation Barrier: The accurate representation of the physical GF MillP800US machine tool as a digital twin within Create MyVirtual Machine was a crucial task for setting up the virtual environment. This involved meticulous geometric modeling, kinematic definition, and ensuring accurate correspondence with the physical machine's behavior. Any inaccuracies in the digital twin could compromise the reliability of simulation results.

2.a.2.2 Configure to order process

This section outlines a Configure to Order process for drone manufacturing, integrating various digital tools and platforms to streamline the workflow from customer configuration to virtual production. The process is designed to ensure consistency and data flow across different stages, as illustrated in the provided *Figure 10*.

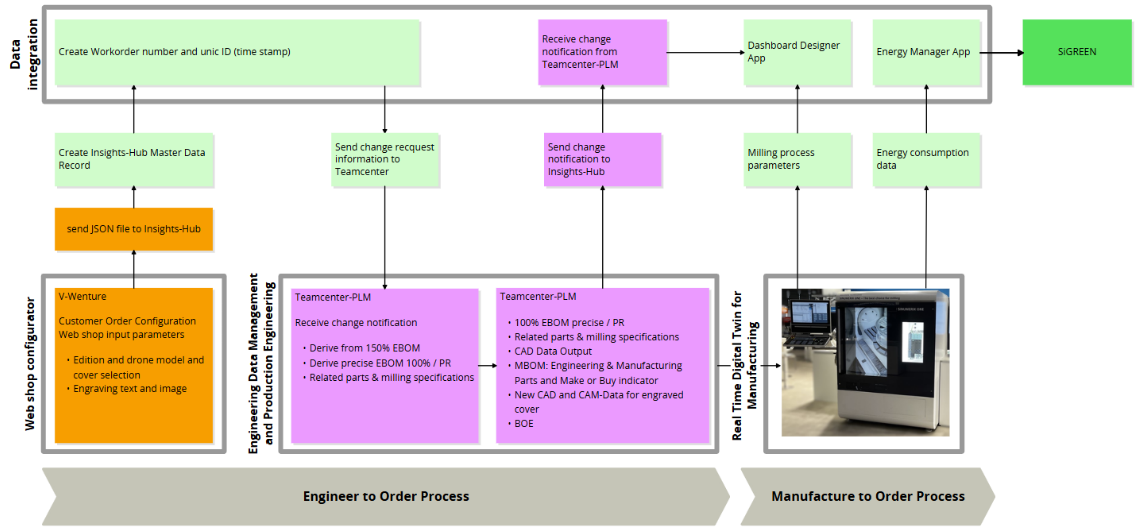


Figure 10 – Configure to order process

The integrated process can be broken down into the following key steps:

- Web shop configurator: A customer initiates the process by configuring a drone within a web shop interface. This configuration includes selecting drone-specific components and defining custom text for engraving on the drone cover.
- Data Transmission to Data Integration Layer: Information regarding the configured drone and the engraving text is transmitted to a central data integration layer, specifically Siemens Insights Hub. Data exchange at this stage occurs through JSON files, ensuring a standardized and efficient transfer of information.
- Master Data Record Creation: Upon receiving the configuration data, Siemens Insights Hub creates a unique master data record. This record is assigned a unique identification number and contains all pertinent drone configuration details.
- Product Lifecycle Management (PLM) Integration: The engraving text and other relevant configuration details are then sent to a Product Lifecycle Management (PLM) system, Siemens Teamcenter. Within Teamcenter, a new Engineering Bill of Materials (EBOM) is derived, incorporating all related specifications for the customized drone and its cover.
- Real-Time Digital Twin: Relevant information, including the updated EBOM and CAD data output, is sent back to the data integration layer as a change notification. Concurrently, this information is also transmitted to a Real-Time Digital Twin model that represents the virtual manufacturing environment. This digital twin, which incorporates tools like Run MyVirtual Machine, virtually produces the drone cover with the specified engraved text. Crucially, this virtual manufacturing process allows for the extraction of valuable data, such as energy consumption and processing time, providing insights into the efficiency and feasibility of the production.

2.a.3 Final KPIs monitoring and validation

2.a.3.1 Industrial Outcomes and Lessons Learned

The FRAISA ToolExpert is fully integrated into Siemens NX CAM software. This allows customers to select the appropriate FRAISA tool directly within CAM and seamlessly transfer the corresponding geometry and cutting data quickly and without errors. As a result, careless mistakes are avoided, saving customers both time and resources.

2.a.3.2 KPI Measurement and Performance Evaluation

Table 1 – KPIs identified for BP1

ID	BUSINESS Indicators List the Business objectives expected for the Business	DESCRIPTION Give a detailed description of the indicators	Unit*	Initial value	M36 Value	Expected final Value	Expect. Date of achievement**

	Scenario/Use Case						
1	Tool selection	Operator takes the right tool and chooses the right strategy	Time to select tool and strategy (min)	60	10	2	2028
2	ToolExpert in CAM	Transfer geometrical tool data and cutting data to CAM	failures (%)	20%	2%	2%	2025
3	Virtual environment for tool and strategy	Programming/set-up time of work pieces can be optimized	Programming/set-up time (%)	100% (base line 120 min)	80%	30%	2028
4	Virtual environment for energy optimization	Estimated energy consumption can be optimized based on tool selection and milling strategy	Energy consumption (%)	100% (base line 80 KW)	80%	60%	2028

2.a.3.3 Final KPI Assessment and Business Impact

Tool selection: Once the KPI tool selection is implemented, FRAISA customers will always use the right tool with the correct cutting data. This strengthens application support and customer loyalty.

ToolExpert in CAM: Largely already implemented. FRAISA thus strengthens customer relationships by enabling NX CAM users to integrate tools more quickly, saving both time and resources.

Virtual environment for tool and strategy: With this KPI, customers reduce setup times, allowing machines to resume production faster and operate more productively.

Virtual environment for energy optimization: Optimally applied tools and strategies reduce customers' energy consumption, enabling them to operate more economically.

2.b Business Scenario 2:

The second business scenario focuses on the tool lifetime, gathering relevant data from the machine and using AI models for predicting the wear of tools. The approach benefits from the federative learning method, collecting data from different machines and aggregating this data for improving the AI models and predictions. Two types of AI applications have been developed by the partners CORE and Atlantis.

2.b.1 Full-scale implementation

2.b.1.1 Architecture

Atlantis Federated Predictive Maintenance (FPdM)

The architecture specifically designed for GF Pilot is illustrated in Figure 11. In this section, components developed by Atlantis for the federated learning objectives are coloured in orange, to provide more clarity of the system's architecture. The Federated Predictive Maintenance (FPdM), corresponds to components number four of the Reference Architecture. The designed system supports two basic workflows:

- Training a model in a federated learning environment
- Performing inference using the trained model on input data

These two workflows have been developing simultaneously during the trials, constantly fine-tuning and improving.

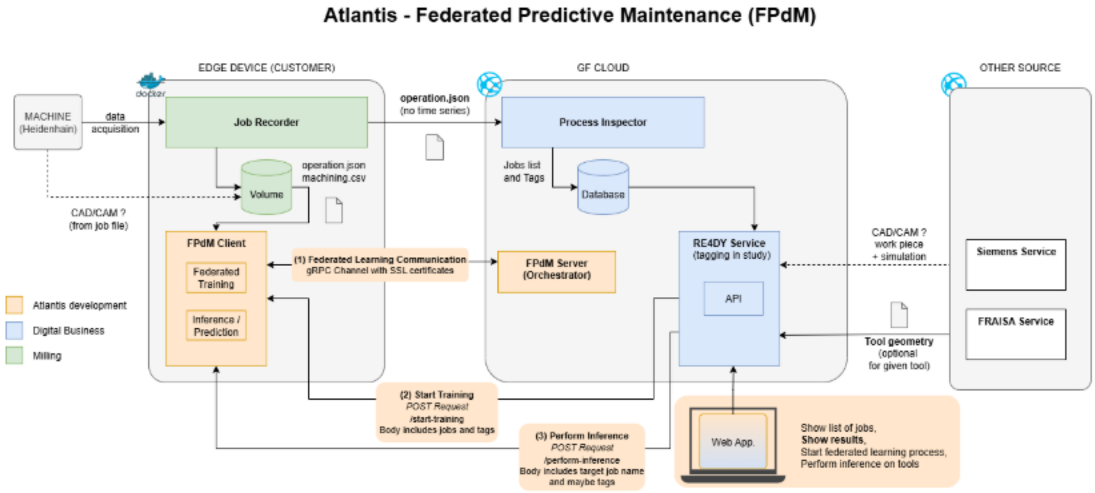


Figure 11 – FPdM architecture

A more detailed view of the federated learning-specific architecture is presented in Figure 12. This schema is adapted from the official Flower documentation, with additional modifications to incorporate a custom Flask Server developed for this use case's requirements.

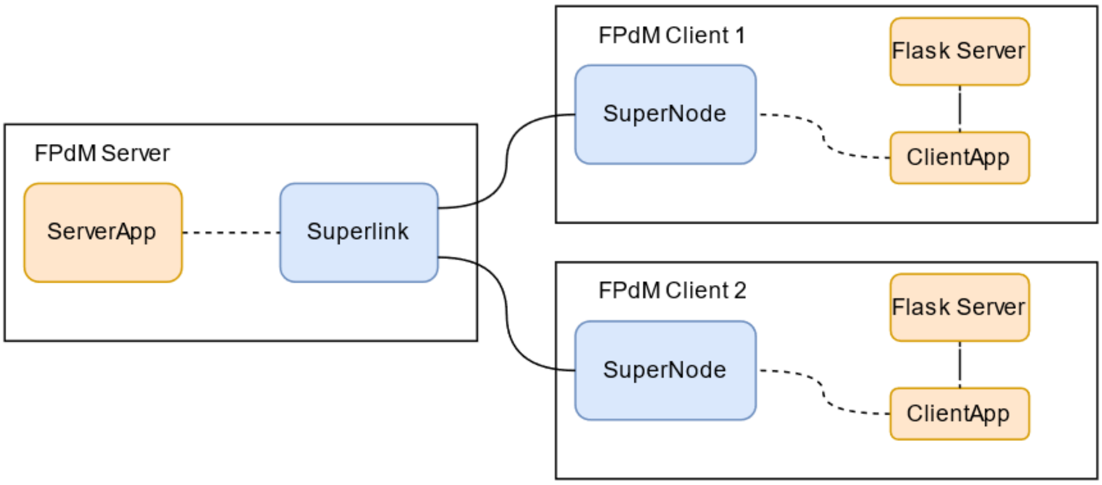


Figure 12 – FPdM Server & Client services

Details for the designed components are described below, as fetched and furtherly enhanced from the official Flower documentation[1]:

- FPdM Server (Orchestrator): This component plays a central role in managing the federated learning process. It handles communication with the participating Clients, orchestrates training rounds, and aggregates metadata received from each client. Based on literature, it should be deployed in a centralized machine that is accessible from all machines that will participate in the federated learning process. The FPdM Server runs two main services:
    - Superlink: This service, included in the Flower federated learning framework, is a long running process that forwards task instructions to Clients and receives task results back. For enhanced security and reduced risks, this service can be configured to accept only Clients that presenting specific SSL (Secure Sockets Layer) certificates during the Client-Server handshake process. Communication between Superlink and any other service (e.g. Serverapp, Supernode) is achieved over gRPC channels.
    - Serverapp: This is a short-lived process with RE4DY-specific code that customizes all Server-side aspects of federated learning systems (client selection, client configuration, minimum required Clients, result aggregation). From this service, the AI engineers can customize, modify or extend the aggregation methods to meet use-case-specific goal.

The Superlink image can be simply pulled from Flower Docker registry while the Serverapp image, that has been designed and built specifically for the RE4DY project, is hosted in the Atlantis repository and can be accessed upon request. Both services can be deployed using Atlantis-provided Docker Compose files, documentation and deployment scripts designed for smooth installation and easy deployment.

- FPdM Client: The FPdM Client is a suite of services designed to run on the edge of milling machines. It consists of the following services:

[1] https://flower.ai/docs/framework/explanation-flower-architecture.html

- Supernode: This service, included in the flower federated learning framework, is a long-running process that connects to the Superlink, asks for tasks, executes tasks and returns task results back to the Superlink.
- Clientapp: This is a short-lived process with project specific code that customizes all client-side aspects of federated learning systems. From this service, developers and AI engineers can define custom data preprocessing methods, specific machine learning models, evaluation and fit functions, and postprocessing methods. This service hosts the custom model that was designed and developed for GF Pilot's use-case. The transmission of any metadata generated by Clientapp to the FPdM Server is achieved via Supernode-Superlink communication.

On the client site, a Flask Server is deployed to facilitate external interactions through API calls. Developed after GF technical team's instructions, the following endpoints are available:

/start-training: Accepts POST requests to initiate and trigger a federated learning session. Upon receiving request, the short-lived Clientapp process is launched using the defined parametrization. The Clientapp process will terminate when all training rounds are completed. The current approach allows dynamic client participation, allowing Clients to join the federation even after training has started. It should be noted that the POST request to the /start-training endpoint, must include milling job names and tags in its body to determine the jobs whose data will be used as training data for the model.

Request Payload Example


```
[{
    "JobId": "3b379f00-132f-4989-85cb-fb9cdf0fa05a",
    "JobName": "2025-05-12T00:00:00.000Z Test Job 1",
    "StartTime": "2025-05-12T00:00:00.000Z ",
    "EndTime": "2025-05-12T00:01:00.000Z ",
    "Tags": [
      "100000000000",
      "Material 1",
      "ap 0.0",
      "ae 0.0",
      "Article Type 1",
      "Vb 0.02",
      "Vbmax 0.00"
    ],
    "Operations": [
      "1a21858e-7ba2-40f4-a7ce-ec7a05fc90eb",
      "0d016e93-9089-4b90-b367-f7dda0eb629f",
      "85026ce5-c6bb-4bc1-be6f-0e1619f3d4c0",
     ]
},
{
    "JobId": "71fdb5ea-5480-4e0d-9f3d-1fa7cd68bcdd",
    "JobName": "2025-05-12T00:01:30.000Z Test Job 2",
    "StartTime": "2025-05-12T00:01:30.000Z ",
    "EndTime": "2025-05-12T00:02:00.000Z ",
```

```
        "Tags": [
          "100000000001",
          "Material 1",
          "ap 0.0",
          "ae 0.0",
          "Article Type 1",
          "Vb 0.025",
          "Vbmax 0.00"
        ],
        "Operations": [
          "e6dfff08-205c-4b90-b00d-35a1cb6d1048",
          "73568d57-5983-4fcb-8ba4-1822a22e5b16"
         ]
    }]
```


Response Example


```
{
      “job_id”:             “038ae0e4-3595-4cd6-9d40-
1cb3d1136920”,
      “message”: “Training started successfully.”.
      “status”: “success”
}
```


/perform-inference: Accepts GET requests and returns the inference made by the trained model. When a GET request arrives, the trained model will be called with the data included in the request given as inputs. The inference result will be returned after the successful data preprocessing and model execution.

Request Payload Example


```
{
“JobNames”:
["2025-05-12T00:00:00.000Z Test Job 1",
"2025-05-12T00:01:30.000Z Test Job 2"]
}
```


Response Example


```
{
    "message": "Inference completed successfully.",
    "results": [
        {
            "JobName": "2025-05-12T00:00:00.000Z Test
Job 1",
            "RUL": 0.6365838646888733,
            "Vb": 0.12731677293777466,
            "timestamp": "2025-05-12 08:40:33"
        },
        {
            "JobName": "2025-05-12T00:00:00.000Z Test
Job 2",
            "RUL": 0.4732838646888733,
            "Vb": 0.0946577293777466,
            "timestamp": "2025-05-12 08:40:33"
```

```
                }
            ]
        }
```


Visualization: To facilitate the visualization of inference results, a dedicated component was developed that allows users to log in and view the latest outcomes for their respective tools. This functionality is implemented using Grafana[2], an open-source platform known for its rich visualization capabilities. After deploying a Grafana instance, a straightforward dashboard was configured, illustrated in Figure 3 and Figure 4. The user can select whether they want to view the dashboard for a single milling tool or for multiple tools simultaneously.

The FEDMA system is designed to enable Federated Learning (FL) in industrial environments while preserving data privacy. It ensures that raw machining data remains at the edge device on FRAISA premises, while the GF Cloud handles model aggregation and orchestration. A web-based UI developed by CORE provides user interaction via GF's RE4DY API.

At the edge, the Job Recorder collects sensor data and metadata during machining jobs, storing them in structured folders (CSV + JSON). The FEDMA Client, developed using FastAPI, exposes RESTful endpoints (/perform-training and /perform-inference) to process these jobs. In a training workflow, the client loads locally labeled data, trains a model, and sends updated weights via gRPC to the FEDMA Server, built with the Flower framework. It then receives and stores the new global model. In inference, the client uses the latest model to predict tool wear and RUL, returning results to the cloud.

The GF Cloud hosts three main components. The Process Inspector receives metadata and human-provided wear labels (Vb, Vbmax). The RE4DY Service serves as an orchestration API, triggering model training/inference at the edge via HTTP. It aggregates predictions and returns them to the user-facing dashboard. The FEDMA Server aggregates updates from all clients (currently one), maintaining the global model using Flower's built-in coordination logic.

Since edge devices are on a secured network, the RE4DY Service acts as an intermediary, relaying API calls to the edge. The UI, hosted by CORE, interacts only with RE4DY, enabling users to trigger jobs and view results. This three-tiered structure—edge, cloud, UI—enables a secure, scalable, and privacy-preserving federated learning workflow.

[2] https://grafana.com/

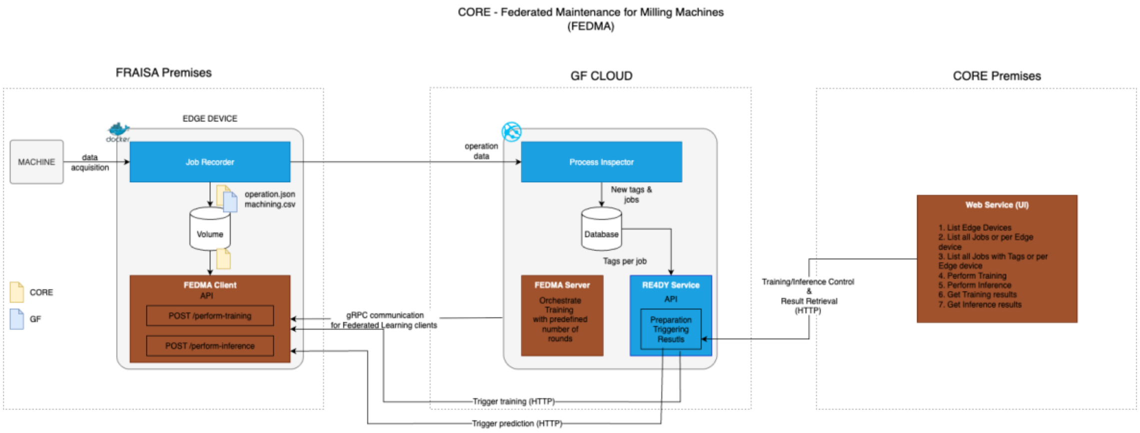


Figure 13 – Federated Maintenance for Milling Machines (FEDMA)

To integrate FEDMA into the GF Cloud, CORE first containerized the FEDMA Client and FEDMA Server, pushed them to GF's Azure registry, and provided deployment guides. Using Docker Compose, GF deployed the components: the server in the cloud, and the client on the FRAISA edge machine with proper volume and network setup.

The FEDMA Server was configured to listen for incoming gRPC connections, while the FEDMA Client was prepared to expose REST endpoints and participate in FL rounds. TLS was enabled for secure communication.

Next, integration testing was conducted. The RE4DY Service (developed by GF) successfully triggered training and inference operations via API calls to the FEDMA Client. The system was verified end to end: local training was initiated, model updates were aggregated by the server, and inference results were returned and visualized.

Since direct access to the edge was not permitted, GF's RE4DY API handled all communication. CORE also developed a UI that interacts solely with RE4DY, abstracting the complexity of the edge operations and allowing users to trigger jobs, view predictions, and input wear labels. This architecture respects security boundaries while enabling robust interaction and federated learning capabilities.

2.b.1.2 AI Models

2.b.1.2.1 FPdM (No 4) Model

The second stage of the workflow focuses on the development and deployment of an artificial-intelligence–based system for predictive monitoring of tool health and wear. For this purpose, a feed-forward neural network (FNN) architecture was conceived, trained, and implemented to address the specific requirements of tool-wear estimation. The design process was guided by both the operational context of the machining environment and the availability of relevant data streams. Two principal categories of input data feed the model:

- Dynamic sensorial measurements, continuously collected and transmitted through the MyRConnect infrastructure. These high-frequency signals capture the real operating conditions of the machine and include, among others, spindle

speed, spindle vibration, spindle load, and curve abscissa. Such parameters are directly related to the physical stresses experienced by the cutting tool and therefore provide essential information for characterising wear mechanisms.
- Static job-related information, describing the machining task as planned and executed. This set comprises features such as total job duration, number of operations, and cutting strategy parameters (for example axial and radial infeed depths). These variables reflect the overall workload imposed on the tool and complement the time-series signals from the sensors.

The neural network is trained to perform a regression task, with the prediction target derived from a fusion of the V_b and V_bmax fields. These fields contain direct measurements of tool wear expressed in millimetres and are considered reliable indicators of the tool's degradation state. By learning the relationship between the combined input features and these wear measurements, the model outputs a continuous estimate of the expected tool wear at the end of a machining job. To convert this raw regression output into a more actionable indicator, the predicted wear value is subsequently compared to user-defined thresholds established by domain specialists. This post-processing step enables the computation of a Remaining Useful Lifetime (RUL) percentage, which quantifies how much of the tool's service life remains before it reaches a critical wear limit. Such an approach provides operators and maintenance planners with an interpretable, real-time metric that can support proactive decision-making, reduce unplanned downtime, and extend overall tool longevity.

2.b.1.2.2 FEDMA AI Models: Federated Maintenance for Milling Machines (CORE) AI Models

At the core of the FEDMA service there is a Deep Learning (DL) model developed to estimate the Remaining Useful Life (RUL) of milling tools. This model enables manufacturers to maximize tool usage, reduce unplanned downtime, prevent failures, and improve equipment reliability.

The model is specifically trained to predict tool wear, using data collected from FRAISA's milling tool experiments and metadata from My rConnect. It leverages both experimental and real-world operational data, including:

- Sensor values: temperature, vibration, cutting forces, etc.
- Machining parameters: axial depth (ap), radial depth (ae), machining strategies
- Tool metadata: article number, capturing the geometry and specification of each tool
- Material properties: such as the type and grade (e.g., M 1.2738 HH)

Wear labels used for supervision:

- Vb: mean wear
- Vbmax: maximum wear

In addition to predicting wear (Vb and Vbmax), the system includes a dedicated RUL estimation algorithm. This algorithm calculates the Remaining Useful Life based on:

- The predicted wear values
- The duration the tool has already been in use
- The tool's geometry and design, inferred from its article number

The model is further improved through a Federated Learning approach, which allows it to continuously learn from new experiments across distributed environments—without centralized data collection—ensuring privacy and scalability.

An inference script integrates the model into operational workflows. Upon receiving a job identifier, the system provides operators with:

- Predicted Vb (mean wear)
- Predicted Vbmax (maximum wear)
- Estimated RUL (remaining useful life)

This allows real-time, data-driven decision-making to optimize tool replacement cycles and improve process efficiency.

2.b.1.3 Applications

The FPdM component is equipped with a high-level user interface that facilitates the interaction with the system and allows users to execute all supported functionalities in an intuitive manner. The interface has been carefully designed to balance analytics visualization and operational actions, providing both insight into tool health and the ability to trigger relevant processes directly from the panel. The interface displayed in the following figure can be conceptually divided into two main sections: the Analytics Section and the Actions Section.

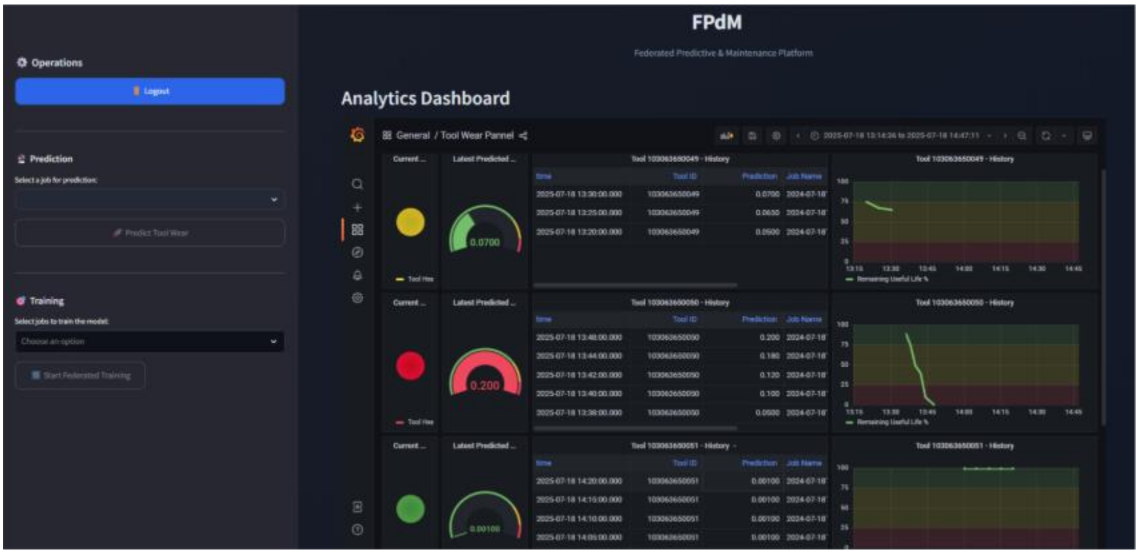


Figure 14 – FPdM User Interface

Located on the right-hand side of the interface, the analytics section hosts a comprehensive dashboard presenting the results of any inferences that have been performed for tools participating in milling jobs. The dashboard is fully interactive, enabling users to select specific tools and define time windows of interest to focus their analysis.

For each selected tool, the dashboard presents four distinct panels arranged horizontally:

Current Health Status Panel: The first panel displays the most recent inference for the tool, reflecting its current health condition. The tool status is visually encoded using a colour scheme, where green indicates low wear, yellow indicates moderate wear, and red corresponds to high wear. This immediate visual feedback allows operators to quickly identify tools that may require attention.

Wear Prediction Panel: The second panel presents the latest wear prediction in millimetres, again using the same colour coding for consistency and quick interpretation. This quantitative metric complements the health status panel by providing precise values for planning maintenance or replacement.

Historical Inference Panel: The third panel contains a table of all historical inferences for the selected tool, enabling users to trace past predictions, monitor trends, and validate model performance over time.

Wear Evolution Graph: The fourth and final panel visualizes the evolution of tool wear over time. This graph allows users to observe progressive degradation patterns, detect anomalies, and correlate tool performance with operational conditions or job characteristics.

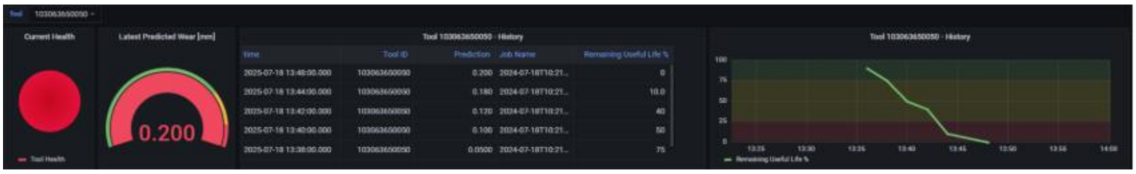


Figure 15 – Analytics section with the designed panels zoomed in for a selected tool

Located on the left-hand side of the interface, the actions section provides interactive fields and buttons that allow the user to execute specific operations. Users can initiate inference for a particular job, by selecting the desired job from the available jobs shown in the list or trigger a federated learning process to update the model with selected jobs forming the training corpus.

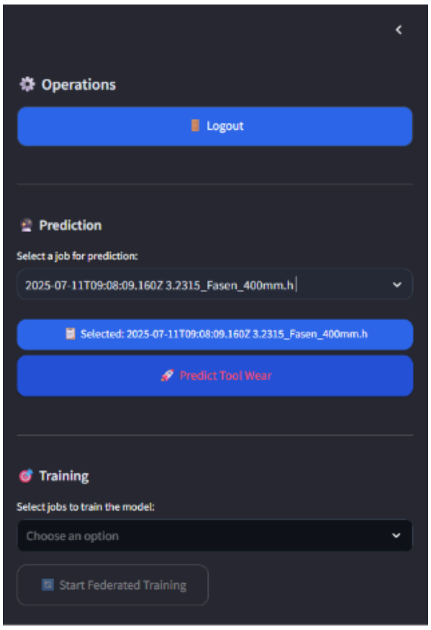


Figure 16 – Select jobs and perform tool wear inference action

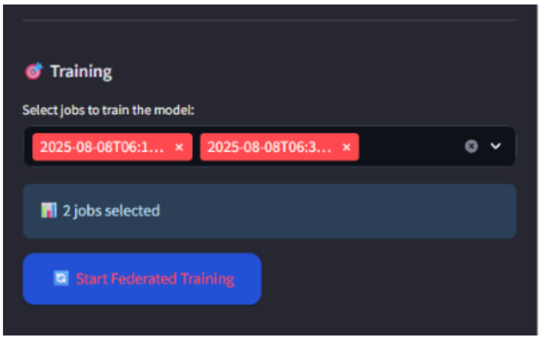


Figure 17 – Select jobs and start training action

By combining both analytics and operational controls within a single interface, FPdM enables a seamless workflow where monitoring, prediction, and model training are tightly integrated. This design ensures that users can not only observe tool health trends but also act proactively, supporting informed decision-making and efficient maintenance planning.

To support real-time decision-making and ensure efficient use of machining resources, CORE has developed an intuitive user interface as part of the FEDMA service. This interface enables operators to interact directly with the federated learning system, providing full control over training schedules and inference execution. By aligning with production workflows, the system empowers users to make data-driven decisions while maintaining operational efficiency and data privacy.

Figure 18 – FEDMA User Interface

Key Functionalities:

1. Controlled Model Training

The interface of Figure 19 allows operators to initiate model training at appropriate times, giving them full control over when training tasks are executed. This ensures that training can be scheduled without disrupting ongoing production and allows optimal use of available resources.

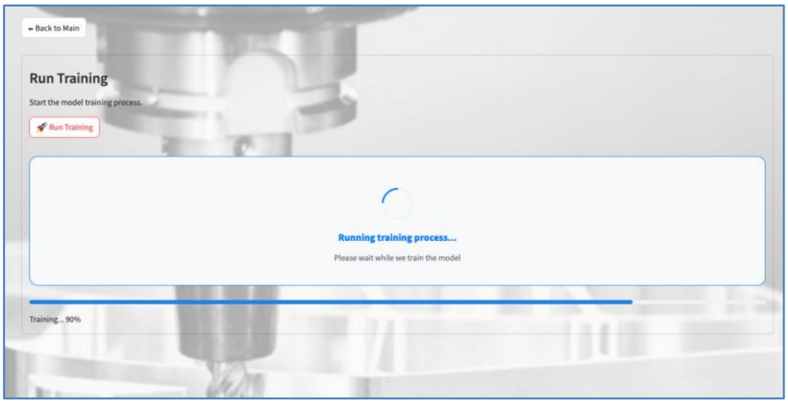


Figure 19 – Initiate FEDMA Model Training through UI

2. Inference Execution for Job-specific Prediction

Through the UI of Figure 20, operators can initiate inference by selecting edge where jobs were executed. This enables users to run predictions on actual completed operations, ensuring that the analysis is both relevant and accurate for the selected context.

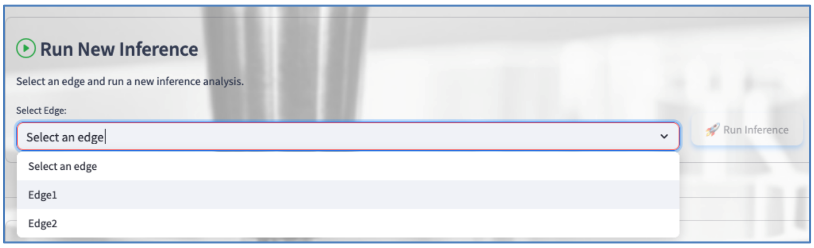


Figure 20 – Initiate FEDMA Inference

Once the inference is executed locally on the edge device, the system generates job-specific predictions, including Vb, Vbmax and RUL estimation, as depicted in Figure 21:

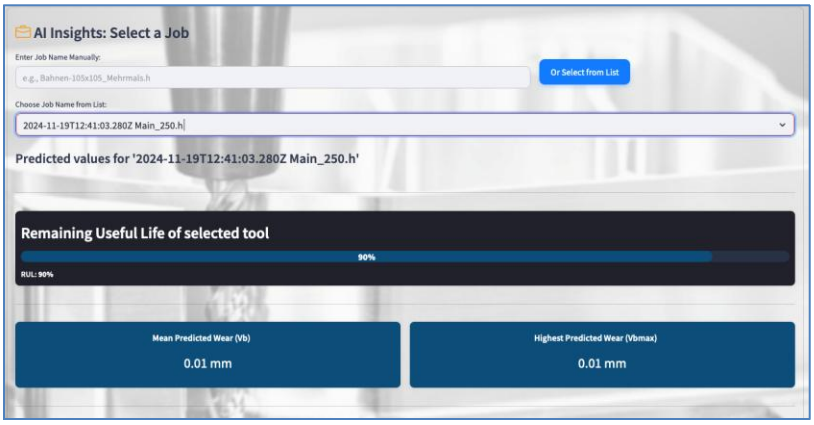


Figure 21 – FEDMA job-specific results

3. User Validation for Continuous Learning

After inference, users can fill in actual wear measurements along with the unique tool ID, as per Figure 22. This feedback loop serves multiple purposes:

- Improves future wear and RUL predictions
- Enhances historical tool tracking and insights
- Supports the model's retraining pipeline within the federated learning framework

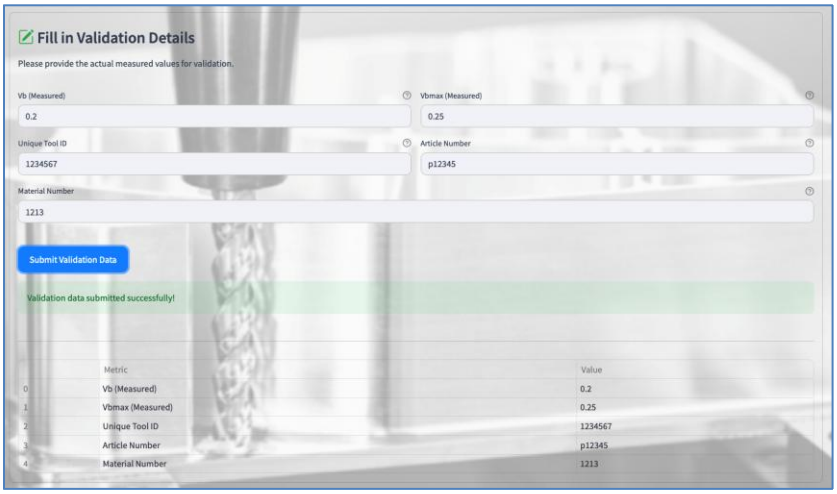


Figure 22 – FEDMA Validation Step

4. Tool Insights and Historical Visualization

After In the Tool Insights section, users can:

- Select a specific tool using its unique identifier
- View wear predictions and RUL estimates from past operations
- Explore wear history through interactive visual plots

FEDMA, through its federated learning approach, depicted in Figure 23, advanced AI models, and user-friendly interface, aims to maximize tool usage, reduce downtime, prevent malfunctions, and enhance equipment robustness—all while respecting data privacy and security. By empowering operators with real-time insights and full control over training and inference, the system delivers a practical and privacy-preserving solution for predictive maintenance in modern machining environments.

Figure 23 – FEDMA tool-specific results

2.b.1.4 Key challenges and solutions for full-scale implementation

The current FPdM (Federated Predictive Maintenance) implementation has successfully demonstrated the feasibility of federated learning by supporting two participating milling machines. However, scaling up to a deployment across a larger number of machines introduces several technical challenges that would need to be addressed to ensure robustness and efficiency.

A primary concern in full-scale implementation is maintaining reliable, low-latency communication between the FPdM Server and all FPdM Clients. As the number of Clients grows, the network traffic will increase resulting to higher demands on bandwidth and synchronization. Unstable or interrupted communication channels may lead to delayed model updates, failed training rounds or inconsistent participation of Clients.

Another critical challenge in full-scale federated learning deployment is managing the computational complexity of both Server and client operations. On the Server side, as the number of participating Clients increases, so does the overhead involved in aggregating model updates, especially when dealing with large models or short training rounds with high-frequency updates. This can strain CPU and memory leading to bottlenecks.

Solutions such as parallel computing or hierarchical aggregation might contribute to addressing these challenges. On the client side, resource limitations are an equally important concern. Edge devices typically come with limited computational power and memory, making it challenging to train complex models. Unlike centralized training approaches, large-scale edge deployments cannot rely on advanced GPUs or high-performance CPUs due to increased costs. As a result, training must be optimized to run efficiently on heterogeneous and resource-constrained hardware. Utilizing adaptive training techniques, such as limiting the number of local epochs, batching strategies or compressing models could help align with computational demands and enable a broader participation of devices in a large-scale deployment.

The full-scale deployment of the FEDMA system presents a range of technical and operational challenges, particularly within heterogeneous industrial environments. Below are the key challenges identified, along with the proposed or implemented solutions, mapped to the core components: Inference, Training, and Federated Learning.

1. Data Integration and Standardization

Challenge:
Ensuring that all required data (sensor operation data, machining parameters, job metadata, and wear labels) is consistently available and structured across different CNC machines and edge devices is a significant integration barrier. Variability in machine software, connectivity protocols, and tag availability introduces risk in system reliability and scalability.

Solution:

- Standardized data format for inference and training input.
- Edge devices validate data (and tags) before triggering training or inference.

2. Computational Constraints at the Edge

Challenge:
Edge devices often lack high-performance computing capabilities, limiting the feasibility of running deep learning models for local training and inference. This is particularly relevant during federated training rounds, which can require significant CPU and memory resources.

Solution:

- Optimized model architecture for edge compatibility (e.g., model pruning, quantization).
- Offer option to operators to start training during machine idle periods, allowing training tasks to run without disrupting ongoing production and making efficient use of available resources.

3. Scalability and Maintainability

Challenge:
As deployments scale to more machines and factories, it becomes challenging to monitor performance, manage updates, and ensure consistent behavior across installations.

Solution:

- Modular software architecture to decouple edge logic, inference, and FL orchestration.

4. Tool-Specific Intelligence and Wear Modelling

Challenge:
Tool behavior varies by geometry and material, and using a generic model may lead to poor performance. Accurate wear prediction and RUL estimation depend on tool-specific behavior and prior usage history.

Solution:

- Use article number as a key input to model tool geometry-specific wear trends.
- Integrate a dedicated RUL estimation algorithm that considers:
    - Predicted wear (Vb and Vbmax)
    - Elapsed tool usage time
    - Tool geometry (from article number)

The FEDMA system has demonstrated functional success across core components: wear prediction, RUL estimation, and federated model improvement. However, scaling to industrial levels requires careful attention to system integration, computational efficiency, network reliability, and test coverage. The outlined solutions serve as a foundation for robust deployment and continued evolution in complex manufacturing environments.

2.b.2 Industrial trials of the pilot

2.b.2.1 Testing procedure and Barriers

For the test and data collection at FRAISA, two machines were equipped with GF My rConnect EDGE boxes, which record process data during milling operations. Each time a program is started on the milling machine, a new job is created in My rConnect. If a tool change occurs during the program, a new sub-job is generated. Tags can be assigned to each job to link additional data. The idea was to use these tags to attach values that are not automatically recorded—such as workpiece material, unique tool ID, tool article number, wear land width, etc.

To make this work, only one tool could be used per job (i.e., per program on the milling machine), because if multiple tools were used, the tags could not be clearly assigned to the individual milling tools.

One of the My rConnect boxes is installed in the R&D test center, where mainly prototype or development tools are used. For these tools, storing data makes little sense because they do not have article numbers, and their geometry data is therefore not stored in the Data Container. The second My rConnect box was installed on a machine where more jobs could be recorded. However, most programs on this machine use multiple tools, which results in sub-jobs that cannot be uniquely assigned.

Attempts were made to generate more data by running separate programs overnight. However, due to the additional effort required for programming, setup, and manually matching the tags, this was only partially successful.

Ideally, tags would be automatically assigned to both jobs and sub-jobs with the start of the program.

2.b.3 Final KPIs monitoring and validation

2.b.3.1 Industrial Outcomes and Lessons Learned

The outcomes and lessons learnt for this business process and scenario are the following:

- Two applications have been deployed by CORE and ATLANTIS for the prediction of lifetime of tools as web applications based on data sharing between Fraisa and GF, using the My rConnect platform.
- The applications use advanced AI models which have been developed through operations data collection campaigns and labelling by Fraisa experts on two GF machines
- The models provide already an accuracy on the prediction of residual lifetime of tools of 80%. It is expected that this accuracy will increase above 90% after deployment and further training used the Federated learning approach, which enables the aggregation of models corresponding to different machines and end user sites, without sharing confidential data. The initial dataset for 1 machine is nearly 100 Machining files, consisting of 5-10 operations, each file around 10 MB)
- The initial objectives have been therefore attained and are technically ready to ramp up in accuracy after deployment, which will enable the achievement of the KPIs related to tooling cost and carbon footprint reduction, delivering a competitive advantage to Fraisa and GF, and enabling a new business model for tooling in collaboration with Core and Atlantis.

2.b.3.2 KPI Measurement and Performance Evaluation

The KPIs for this business process are represented in Table 2. They are focused on tooling cost reduction, enabled by the prediction of lifetime that allows to use the tool closer to its end of life without risks; associated with the increase of tool lifetime KPI. Additionally, the knowledge and understanding of tool lifetime for given applications enable the KPI of better designed tooling and finally the reduction of carbon footprint, which is associated with the material waste of the tool at end of life. Avoiding tool breakage enable the recycling of tools 2-3 times.

The previous KPIs are enabled by the accuracy in the prediction of the tool lifetime through the algorithms and applications developed by Core and Atlantis. Currently this accuracy is of 83%. This enables a mean gain of nearly 15% in tool lifetime, as the mean residual lifetime of tools when they are changed by users, compared to maximum lifetime, is currently 70% and below. This allows to use the tools 15% more time and reduce the costs by around 20%. With the usage and deployment of the applications it is expected to reach a 30% in cost reduction, by extending lifetime by 30%. The rate of recycling of tools will be also facilitated, and it is expected that the carbon footprint will be reduced by 10% in 2026. The lifetime of tools is currently being optimised for special applications based on the knowledge accumulated with the solution, and the increase of this lifetime is of 20%, as depicted by Table 2.

Table 2 – KPIs identified for BP2

ID	BUSINESS Indicators List the Business objectives expected for the Business Scenario/Use Case	DESCRIPTION Give a detailed description of the indicators	Unit*	Initial value	M18 Value	Expected final Value	Expect. Date of achieve ment**
1	Tooling cost reduction	Due to optimized application parameters tool can stay longer in operation	Tool cost (€)%	100%	80%	70%	2025
2	Longer tool life cycle	Increase of tool lifetime	(€)%	100%	120%	130%	2025
3	Better designed tooling	Tool layout is tailored on the application	Life time h (%)	100%	100%	120%	2026
4	Reduced CO2 footprint	Less energy consumption	kW (%)	100%	100%	90%	2026

2.b.3.3 Final KPI Assessment and Business Impact

The final KPIs assessment gives out significant business impact for the organisations involved in the development of this business processes:

- As main stakeholder, Fraisa has now a unique solution for adding value to the tools they manufacture. On one hand they can propose longer lifetimes during usage, currently 15-20%, and expected to increase up to 30% in 2026 following improvement of the accuracy of algorithms. On the other hand they decrease manufacturing costs by recycling efficiently the tools before the break, saving therefore materials and reducing the carbon footprint of manufacturing.

- GF, as machine manufacturer, has a high value technology offer in partnership with Fraisa. It is well known that manufacturing costs can be strongly influenced by tooling costs, in particular in key markets like medtech and aerospace. The increase of tool lifetime has therefore a direct influence on the cost per part, which may be reduced by 10% and provide a competitive advantage with respect to standard tooling systems.
- For CORE and ATLANTIS, in charge of the development of the prediction algorithms and application deployment, the success of the project implies the implementation of a new business model where they can gain benefit from either subscriptions or licensing of the software, with an initial offer giving already significant value to the organisations involved and their customers, and a potential for improvement and greater revenues as the application usage increase and develops with the partnership.

2.c Business Scenario 3:

The third scenario of the pilot focuses on the machine maintenance, starting from the machine spindle, the most critical component, and then extending the scope to the rest of main components subjected to intensive efforts and wear. The resulting applications, VEGA and Machine Diagnosis, incorporate different types of models. A particular attention was made to the drive train, for which a detailed ontology was developed, as well as an AI model for predicting potential failures.

2.c.1 Full-scale implementation

2.c.1.1 Architecture

The architecture of the machine predictive maintenance applications is represented in Figure 24 and Figure 25. A new acceleration sensor of high resolution has been integrated into the spindle and integrated in the connectivity framework of the machine through the EDGE computer and towards My rConnect platform. A visualisation interface, VEGA monitoring and analytics, was developed for displaying the parameters.

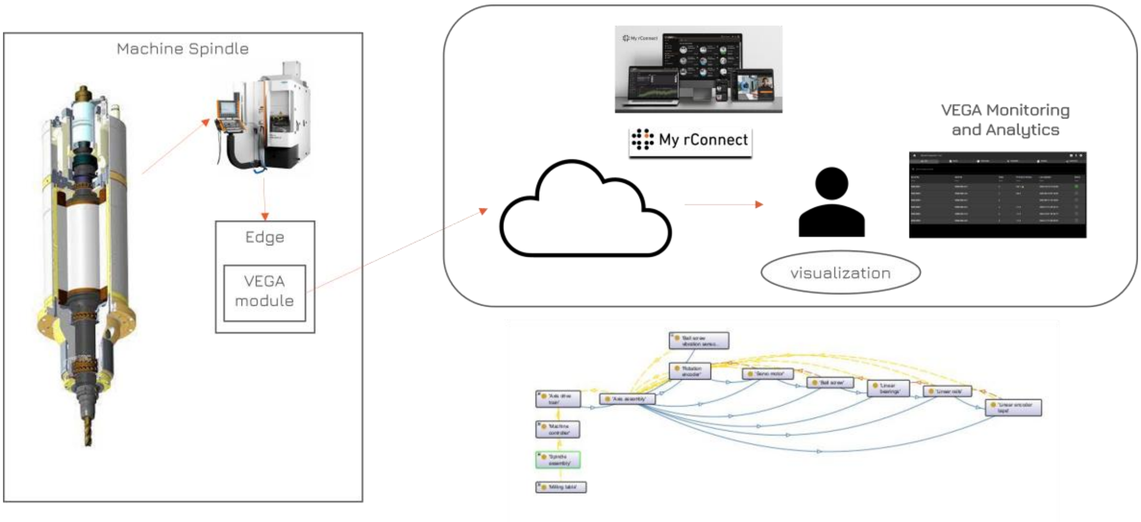


Figure 24 – VEGA Machine Spindle Monitoring Application Architecture

For extending the application to the other critical components, another application was developed, Machine Diagnosis. For the specific case of the drive train, an ontology model was developed by UiO, as well as an AI predictor of potential failures.

A high-level architecture for the Machine care application is represented in Figure 25. Seven critical components were identified, and specific sensors were deployed on My rConnect, The application supports dedicated tests in controlled environments and conditions, where models can provide accurate state of health and predict probability of failures, giving then advice to service technicians for maintenance work, as synthetised by Figure 25.

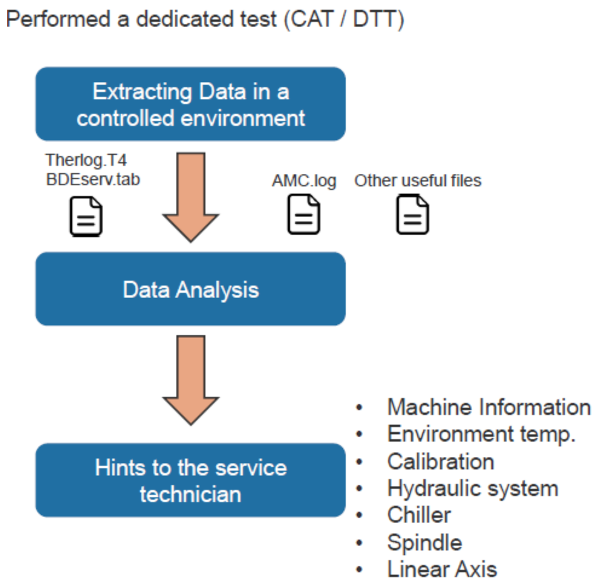


Figure 25 – Machine Diagnosis Application Architecture

2.c.1.2 AI Models

A particular AI model was developed for predicting potential malfunctions or failures for the drive train of milling machines. The solution is based on data collected during production and maintenance tests, using the different available sensors on the machine. A My rConnect application gather this data and a machine learning model was trained using expert diagnosis knowledge from technicians.

When comparing the accuracy of the predictions with expert labelled data we observe that the main critical issues were identified, with none of the critical issues disregarded by the model. This model is also a conservative one as some events classified as critical by the model were not considered the same level by the experts.

The model (whose confusion matrix is reported in Figure 26) is therefore mostly accurate and most importantly does not disregard critical issues and can be used now by technical experts for automating the control of machines during manufacturing and during maintenance tests.

- Confusion Matrix

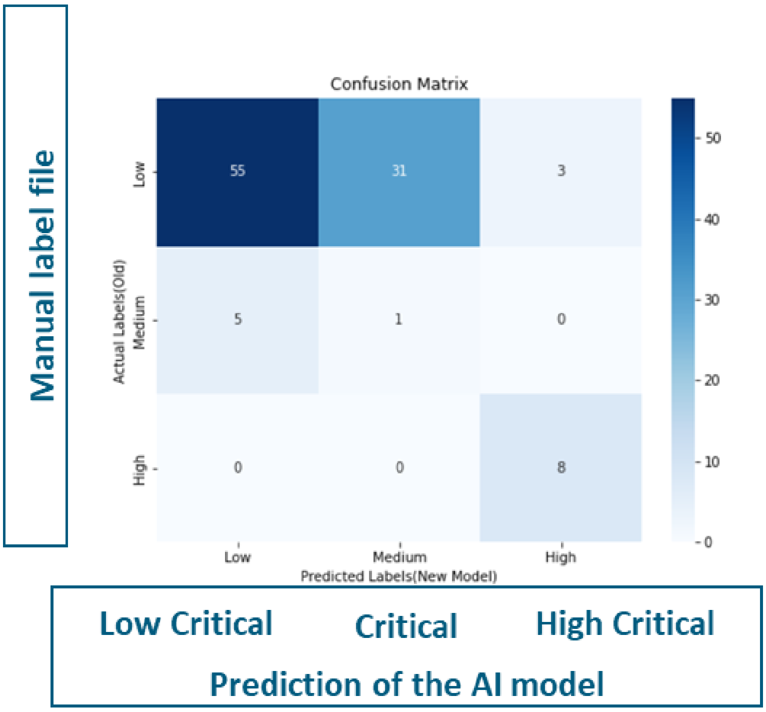


Figure 26 – Confusion matrix for the resulting predictions of the machine learning model for the drive train test, as compared to the expert labelling

2.c.1.3 Applications

The monitoring of critical machine components is made by specific applications in the My rConnect environment. These applications have been developed in two modules; the VEGA monitoring, for the spindle sensors and status, and the Machine Diagnosis application, covering a larger number of critical components of the machine, from the drive train to the chiller and the hydraulic system.

Figure 27 to Figure 29 show the interface of the VEGA spindle diagnosis application. Figure 27 shows the information on the spindle status which can be displayed in the interface, including identification number, axis and motor parameters, operation statistics and main sensor information in real time.

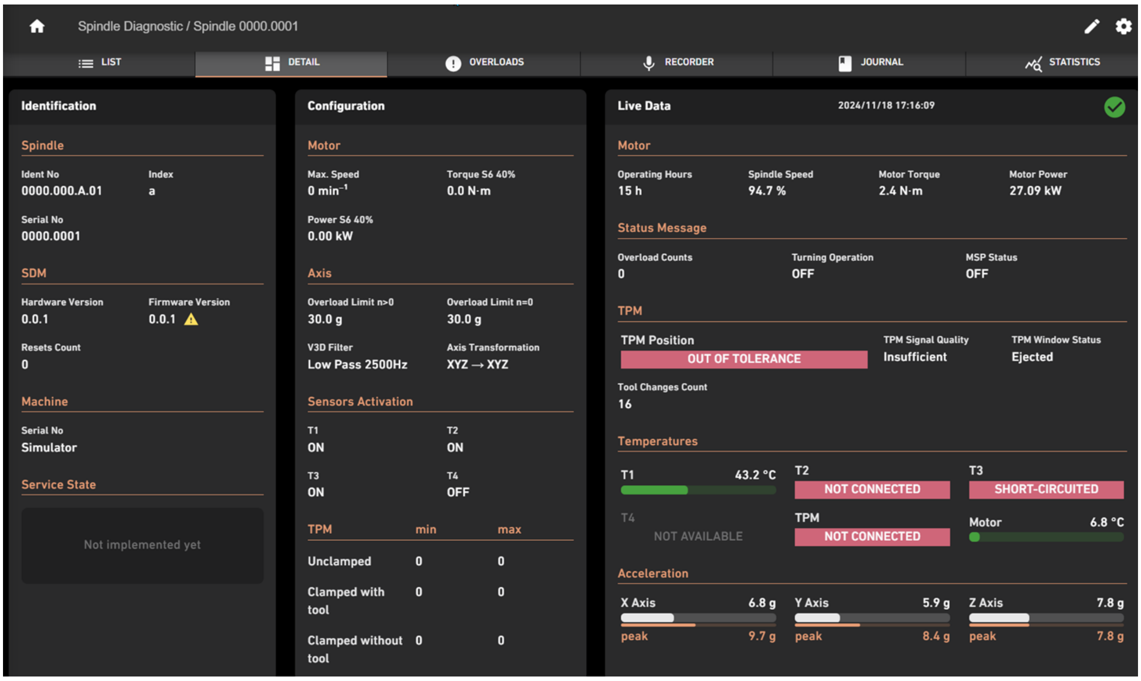


Figure 27 – VEGA Spindle diagnosis interface

The application can go into a more detailed statistics and monitoring information about spindle sensor temperatures, as this is a critical indicator of the quality of the operation and potential failures with the system. This is represented in Figure 28.

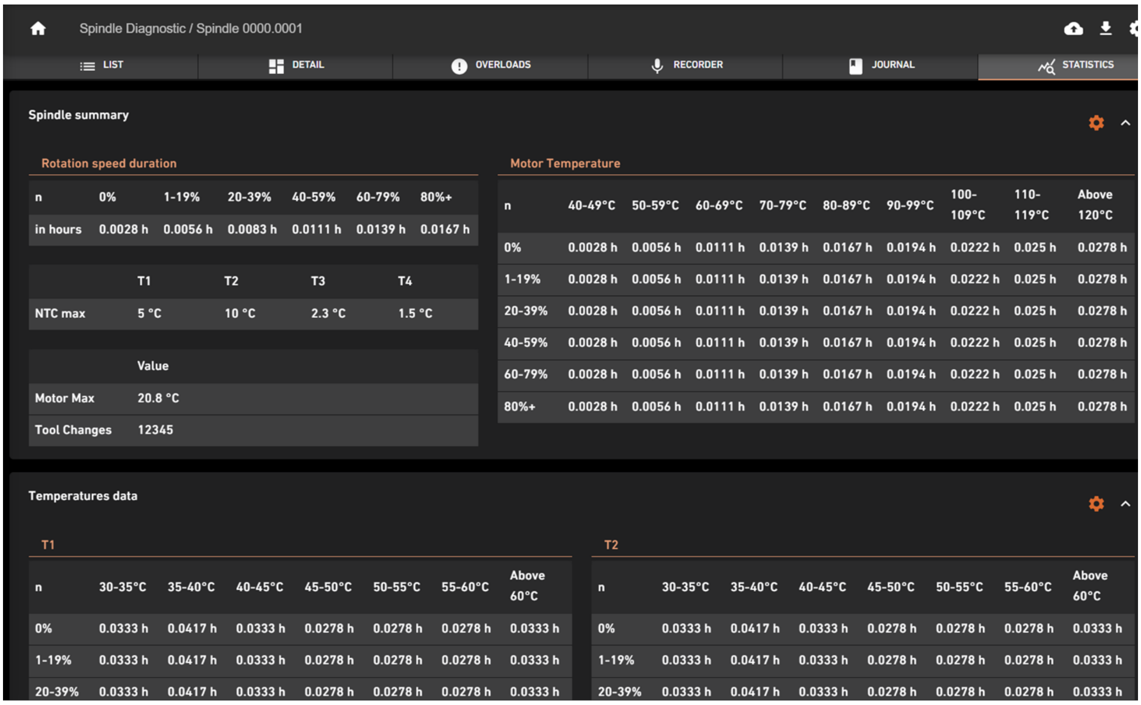
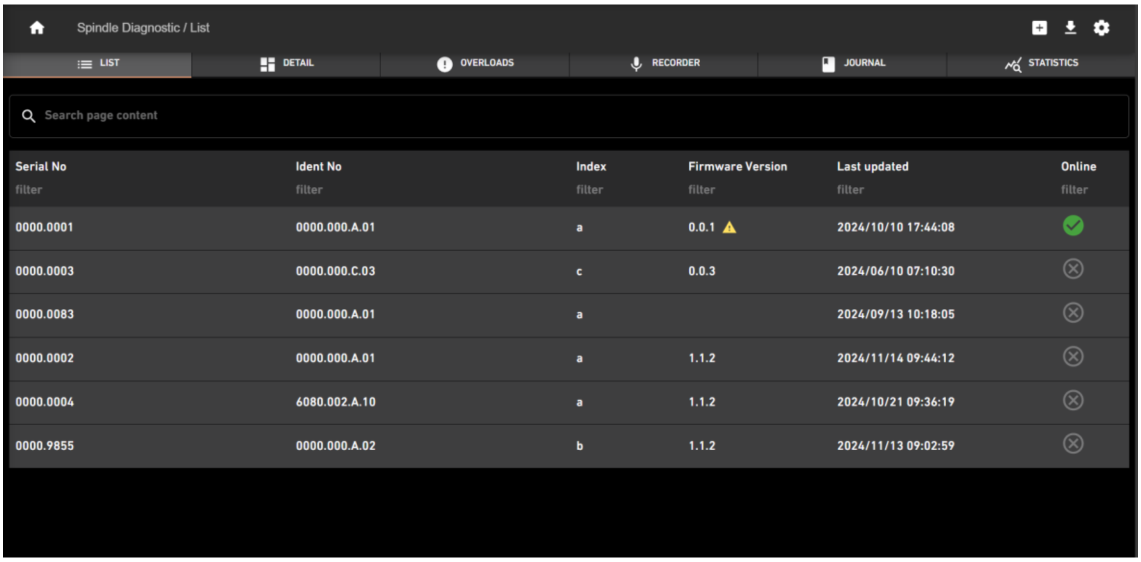


Figure 28 – VEGA Spindle diagnosis sensor statistics

The application allows to gather information from different machines and spindles, verify its status and go into a more detailed assessment of each spindle condition (Figure 29).



Serial No	Ident No	Index	Firmware Version	Last updated	Online
filter	filter	filter	filter	filter	filter
0000.0001	0000.000.A.01	a	0.0.1	2024/10/10 17:44:08	
0000.0003	0000.000.C.03	c	0.0.3	2024/06/10 07:10:30	
0000.0083	0000.000.A.01	a		2024/09/13 10:18:05	
0000.0002	0000.000.A.01	a	1.1.2	2024/11/14 09:44:12	
0000.0004	6080.002.A.10	a	1.1.2	2024/10/21 09:36:19	
0000.9855	0000.000.A.02	b	1.1.2	2024/11/13 09:02:59	

Figure 29 – VEGA Spindle diagnosis for different machines and spindles in a shopfloor

In a second step a second application, Machine Diagnosis, has been developed for covering a larger number of critical components, including the chillers, the axes, the drive train, the hydraulic system and different sensors around the machine for monitoring the environment and mechanics. The application integrates diagnosis algorithms based on data collection and labelling by experts, which enable to identify potential failures of the components and the machine, or identify quality issues on the part. The interface of the application is in Figure 30.

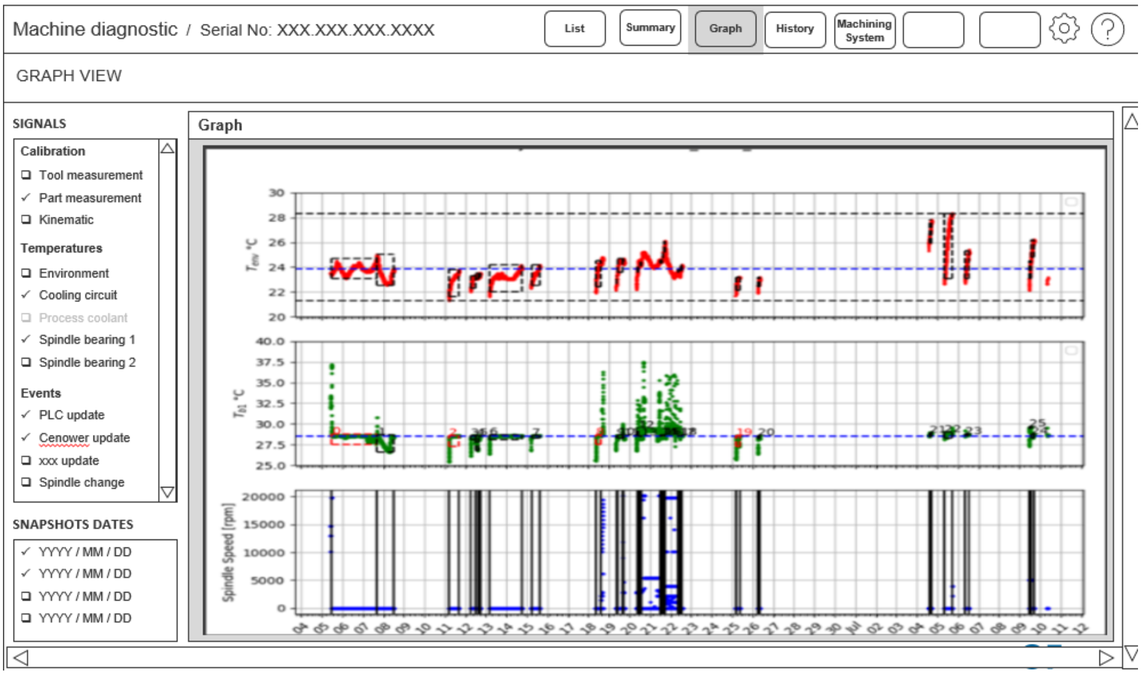


Figure 30 – Machine Diagnosis Application Interface

2.c.1.4 Key challenges and solutions for full-scale implementation

The main challenges for the full-scale implementation of the application are related to the effectiveness of the diagnosis system. The solution needs to be adopted by the customers by gain their trust and helping them to improve the uptimes and reduce costs of maintenance of the machine. Although the algorithms reach an accuracy of 80%, the remained uncertainty can be detrimental to this adoption.

The solution for the challenge is based in a first deployment to be done only at the level of service technicians. This stage will help to correct and improve the features of the application and the accuracy of predictions, so to enable the final fulfilment of the KPIs related to machine uptime and cost reduction. Once this is validated, a second stage will be implemented for deploying the solution at the level of end user customers.

2.c.2 Industrial trials of the pilot

2.c.2.1 Testing procedure and Barriers

The industrial trials of the pilot are made in the following steps:

- Tests at machine production: The system is part of the production line of the machine. The different critical components of the machine are tested using the new applications in order to verify the appropriate assembly and functional performances of the critical components.
- Tests at service level: The system is tested during installation of the machine and during maintenance activities of the Service organisation, by specialised technicians. This phase guarantees performances at installation and supports the service technicians for diagnosing the machine during maintenance activities at the customer site.

There are no barriers for these two phases, except the current potential issues for connecting the application, as a webservice, at the customer site. This is a generic potential issue for the My rConnect platform and its applications, and current procedures guarantee this deployment for different use cases.

2.c.3 Final KPIs monitoring and validation

2.c.3.1 Industrial Outcomes and Lessons Learned

The following outcomes and lessons learned can be drawn from the KPIs evaluation and monitoring for the Machine Diagnostics application:

- The VEGA Spindle monitoring and Machine Diagnostics applications are currently being deployed at internal machine production and services levels.
- The applications will be part of a subscription package providing continuous diagnostics services and preventative maintenance based on the diagnostics recommendations

- The subscription packages will provide benefits regarding machine uptime and maintenance costs with respect to the current service, based on periodic maintenance and replacement of components after failure.
- Both machine uptime and maintenance costs are the most critical factors affecting the production costs of customer parts, and preliminary feedback indicates a great interest by customer for implementing the packages and solution, which also influences the buying decision of the customers
- Marketing activities for the solution will be oriented to the communication of such unique value proposition, guaranteeing uptimes and minimum costs of replacements, avoiding unexpected failures and expensive replacement of critical components.

2.c.3.2 KPI Measurement and Performance Evaluation

The KPIs for this business processed are represented in Table 3. There is a modification of the KPI no. 2, which was previously based on the Remaining Useful Time of key component before refurbishing, which should be minimised in order to reduce costs. This KPI is rather difficult to measure so it has been updated to reflect the machine maintenance costs reduction with the solution as compared to the current preventative situation, having thus a baseline of 0% reduction prior to the installation of the system.

Table 3 – KPIs identified for BP3

ID	BUSINESS Indicators List the Business objectives expected for the Business Scenario/Use Case	DESCRIPTION Give a detailed description of the indicators	Unit*	Initial value	M18 Value	Expected final Value	Expect. Date of achieve ment**
1	Machine Uptime	Productive machine time with respect to total available time	%	80 %	90%	95%	2026
2	Maintenance costs for end user	Cost reduction with respect to standard maintenance costs	%	0%	20%	30%	2026

2.c.3.3 Final KPI Assessment and Business Impact

The solution is currently assessed at development and production levels. The first interesting outcome is the increase of efficiency at production for the control of critical components and the identification of potential defects at assembly stages and final controls. There is also valuable feedback from field tests at customer level made by the company services organisation. In this particular case the estimated gain in machine uptime is of the order of 10% (from a baseline of 80%). The commercial deployment is expected to rise the accuracy of the solution and therefore the machine uptime by an additional 5%, to a maximum expected level of 95% by the customers.

The preliminary feedback of field test customers indicates an improvement as well in the KPI of cost reduction related to maintenance. The current system is based on periodic controls and exchange of critical components based on mean time to repair statistics for each item. This does not take into account the changes in the production system at customer workshop level. The implementation of the new service, based on a subscription and continuous monitoring of the components gives now an estimated gain of 20%. During the deployment at large scale, the expectation is to achieve a cost reduction of 30%. This KPI will be carefully measured taking into account the historic costs of maintenance and replacement of failed components, which represents the most important factor for services expenses for the customers.

2.d Business Scenario 4

This business scenario shifts quality control from a post-process activity to a proactive, in-process system. It focuses on integrating metrology directly into the machine tool to create a closed-loop feedback mechanism. This enables real-time compensation for machine geometry and part-setup errors and allows for dynamic adaptation of the machining program based on in-process measurements, significantly reducing scrap and ensuring final part precision.

2.d.1 Full-scale implementation

The full-scale implementation of the Adaptive Digital Manufacturing scenario was conducted within the industrial environment of the Fraisa facility. The setup was centered around a GF machine tool with a Siemens CNC controller, which was integrated with an advanced in-machine metrology solution.

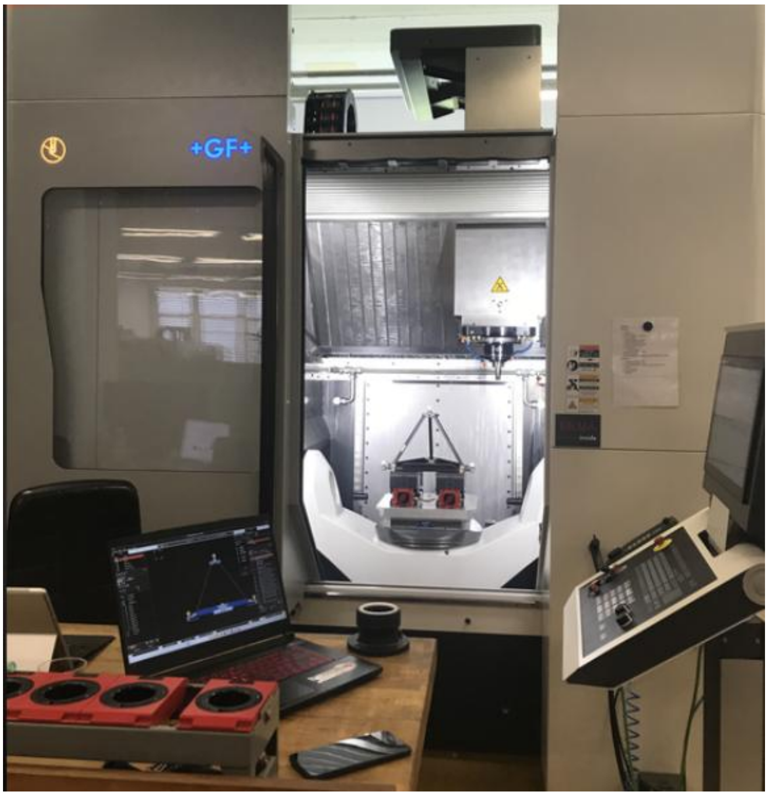


Figure 31 – Business Scenario 4 setup

The physical implementation, as shown in Figure 31, consists of a spindle-mounted touch probe for data acquisition, connected to a dedicated PC running the specialized M3 metrology software, that served as the central edge computing component. It was responsible for orchestrating the measurement cycles and processing the data in real-time. This setup enabled a direct communication link with both the machine's native Siemens CNC controller for closed-loop actions and with the GF myR-connect cloud platform for data aggregation and analytics.

This configuration was designed to validate the complete data flow, from physical measurement on the machine to the execution of adaptive corrections and the subsequent storage of quality data.

2.d.1.1 Architecture

The architecture for the Adaptive Digital Manufacturing scenario is designed as a hybrid system, integrating edge computing components with the machine tool's native controller and a cloud platform. The solution is centered around the M3 Metrology Software, a key component of the RE4DY Toolkit, which manages the in-process metrology operations. In this pilot, the M3 software was deployed on a dedicated PC.

The key components and their interactions are as follows:

- Machine Tool & Sensor: A GF machine tool is equipped with a spindle-mounted touch probe. This probe is the primary data acquisition device, physically interacting with the workpiece and machine artifacts, and feeding raw data to the edge component.
- M3 Metrology Software: The M3 Metrology Software, running on the dedicated PC in this implementation, orchestrates the measurement routines, processes the raw data from the probe, and computes the necessary corrections and results. This software is capable of flexible deployment, potentially on other devices or directly integrated into advanced CNC controllers.
- Control-Loop Integration (Machine Level): The M3 Metrology Software communicates directly with the Siemens machine controller (CNC). It sends critical data for real-time, closed-loop adjustments:
    - Machine Calibration Data: To compensate for geometric errors of the machine, utilizing an artifact for verification.
    - Part Alignment Results: To dynamically adjust the machining program's coordinate system, correcting the part's setup.
- Analytical-Loop Integration (Platform Level): For monitoring, traceability, and higher-level analytics, the M3 Metrology Software sends the final measurement results in the standardized ISO 23952:2020 - Quality information framework (QIF)format to two destinations:
    - The GF myR-connect platform, which serves as the central data aggregator for the pilot.
    - The Siemens controller, for local data logging and quality verification purposes.

This architecture effectively enables a closed-loop control system at the machine level while simultaneously pushing rich, structured quality data to a platform for broader analysis and process insight.

2.d.1.2 Applications

The core application utilized in the Adaptive Digital Manufacturing scenario is the M3 Metrology Software from Datapixel (Figure 32), which serves as both the processing engine and the primary user interface. Deployed on the external Industrial PC, M3 is a comprehensive metrology software solution that enables several critical functionalities:

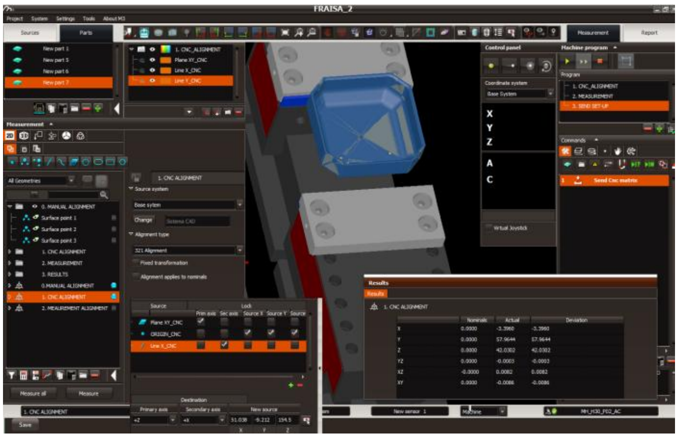


Figure 32 – Part measurement in the M3 software

- Machine Calibration: The software allows for the verification and calibration of the machine tool's geometric accuracy using specialized artifacts. This process is fundamental to compensate for positioning errors and ensure the machine performs within specifications.
- Part Alignment: M3 is used to measure the exact position and orientation of the workpiece once it is fixtured on the machine. It then calculates the necessary coordinate system adjustments to align the theoretical machining program with the actual part setup, a key step in adaptive manufacturing.
- Measurement Program Development: Engineers use M3 to create and define complex measurement routines, specifying probe paths, feature definitions, and analysis parameters.
- Measurement Execution: Operators can initiate and monitor in-process measurements directly through the M3 interface, ensuring the correct execution of the defined programs.
- Measurement Simulation: The software provides simulation capabilities, allowing for the validation of measurement programs offline before their deployment on the machine, optimizing efficiency and preventing potential collisions.
- Data Visualization and Analysis: M3 offers tools for visualizing measurement results, analyzing deviations, and generating reports, which are crucial for understanding the part's quality and the machine's performance.

While the Siemens CNC controller acts as the recipient of correction data and the myR-connect platform aggregates QIF results for broader analytics, the M3 Metrology Software is the direct application interface that facilitates all the key metrological tasks within the pilot.

2.d.1.3 Key challenges and solutions for full-scale implementation

Challenge: Limited Industrial Asset Availability

The full-scale testing of the Adaptive Digital Manufacturing scenario was constrained by the limited availability of the designated industrial asset. The metrology software and sensor were integrated into a specific GF machine at the Fraisa facility, which was under a time-limited lease. This lease expired before the complete test plan could be executed, posing a significant risk to the pilot's validation activities.

Solution and Mitigation Strategy To mitigate this, a short-term agreement between GF and Siemens was secured, granting a brief extension. This window was sufficient to conduct one full round of testing, which successfully demonstrated the core capability of the architecture: metrology data was correctly extracted from the process and integrated with other data streams. To complete the pilot's objectives, further validation activities were carried out using the metrology equipment located at SSF (Figure 33). This equipment features a similar setup to the one at the Fraisa facility, ensuring the comparability and relevance of the results.

Figure 33 – Metrology Equipment at the SSF used for the final tests

2.d.2 Industrial trials of the pilot

2.d.2.1 Testing procedure and Barriers

The industrial trials were structured to validate the three core capabilities of the Adaptive Digital Manufacturing scenario as distinct components. Each component was tested individually to confirm its functionality and integration with the pilot hardware and software.

- Machine Verification Trial (BP4-C1): This test focused on the ability to measure and compensate for the machine's geometric errors, both linear and rotatory axis related (see Figure 34). The procedure involved using calibrated artifacts (a tetrahedron and a sphere) placed within the machine's working volume. The M3 software executed a measurement routine to capture the artifact's geometry, calculated the machine's kinematic errors (e.g., perpendicularity, positioning), and generated the corresponding compensation parameters for the controller. The objective was to verify the system's capability to perform a machine health check and calibration automatically.

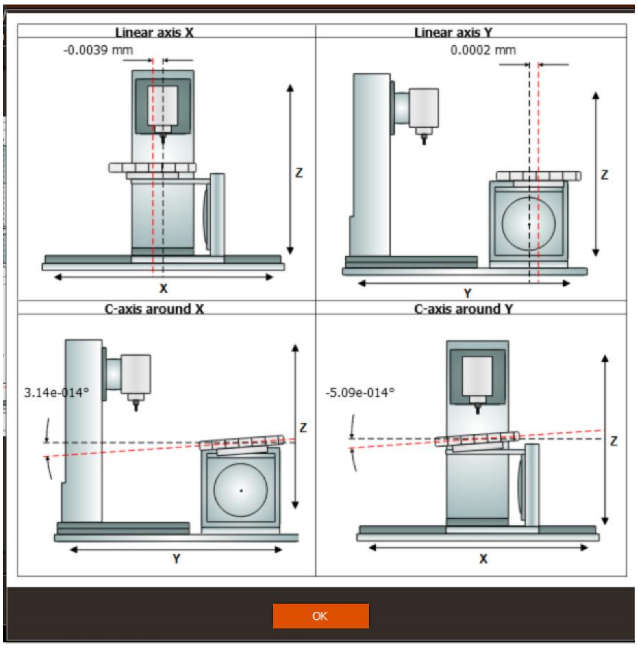


Figure 34 – Multi-axis machine calibration

- Adaptive Part Alignment Trial (BP4-C2): This trial validated the system's ability to correct for part setup inaccuracies. A workpiece was fixtured in the machine, and a measurement program based on its CAD model was run to identify its actual position and orientation. The trial successfully demonstrated that the M3 software could compute a roto-translation matrix and establish a new, corrected coordinate system, ensuring that the subsequent machining operations would be perfectly aligned with the part.

- Automated Metrology & Feedback Trial (BP4-C3): This test demonstrated the in-process quality control capabilities. The trial consisted of executing a measurement program on a machined part to verify critical features against their CAD specifications. The key outcome was the successful generation of a measurement report in the standardized QIF format, confirming that the system could extract quality data and share it with other platforms like myR-connect for analysis and traceability.

2.d.3 Final KPIs monitoring and validation

2.d.3.1 Industrial Outcomes and Lessons Learned

The outcomes of the Adaptive Digital Manufacturing pilot demonstrate a clear progression from foundational work and simulation to physical implementation and validation. The lessons learned reflect the maturity gained throughout this process.

Phase 1: Foundational Work and Simulation

The initial phase of the project focused on establishing the necessary groundwork in a controlled, offline environment. The key outcomes from this stage were:

- Definition of Data Exchange Standards: The data formats and schemas required for all adaptive processes were specified, with a focus on adopting the QIF standard for interoperability.
- Offline Process Validation: The core functionalities of part alignment and in-process measurement were developed and validated through simulation using CAD-based measurement programs. This allowed for the refinement of the logic without consuming machine time.
- Development of Integration Components: The M3MH postprocessor, the software component required for the M3 software to communicate with the Siemens controller, was specified and implemented, preparing the ground for physical integration.

Phase 2: Physical Implementation and Validation

Building upon the foundational work, the next phase involved deploying and testing the solution in the industrial setting at the Fraisa facility. The main outcomes were:

- Successful On-site Deployment: The M3 Metrology Software was successfully installed and proven to be fully functional on the target Siemens machine.
- Demonstration of Core Capabilities: The three key capabilities were executed successfully on the machine:
    - Machine verification and calibration using a calibrated artifact.
    - Automated part alignment based on CAD data.
    - In-process quality measurement and the generation of standardized QIF data files.
- Data Sharing and Integration: The pilot demonstrated the ability to share the generated QIF data through a Data Space connector, using the Innovalia Data Space infrastructure.

Key Lessons Learned

- The "Simulate First" approach is highly effective: The initial focus on simulation was crucial. It enabled a faster and lower-risk deployment in the physical phase, as most of the process logic was already validated.
- Standardized data formats are essential for interoperability: The early definition and subsequent implementation of the QIF format were key to ensuring that the quality data was ready to be shared and consumed by other systems, like a data space.
- Logistical planning is as critical as technical development: The primary lesson learned, reinforced by the project's challenges, is the critical importance of securing long-term access to industrial assets for the final stages of process evaluation and KPI validation.

2.d.3.2 KPI Measurement and Performance Evaluation

The performance of the pilot was evaluated against three specific business indicators identified at the project's outset. These KPIs focus on improvements in speed, quality, and overall efficiency. The verification method for these KPIs involves comparing the performance of the new automated process against the traditional, manual baseline.

Table 4 summarizes the expected performance for each KPI.

Table 4 – KPIs identified for BP4

ID	BUSINESS Indicators List the Business objectives expected for the Business Scenario/Use Case	DESCRIPTION Give a detailed description of the indicators	Unit*	Initial value	M40 Value	Expected final Value	Expect. Date of achievement**
1	Machine Verification Time	Time required to perform a full machine verification using an artifact.	H	8	2	2	2025
2	Production Scrap Rate	Percentage of non-conforming parts due to machining errors.	%	5%	3%	1%	1 year after the project
3	Production Cycle Time	Overall time from raw part setup to finished part.	Min	120	115	108	1 year after the project

Note on Verification: While the industrial trials successfully demonstrated the technical functionalities required to achieve these KPIs, the limited machine availability prevented a long-term statistical validation. The "Expected final Value" reflects the targets based on the successful execution of the automated routines in the controlled tests. The initial values are representative examples of a traditional manufacturing baseline.

2.d.3.3 Final KPI Assessment and Business Impact

The achievement of the defined KPIs through the RE4DY solution provides a significant and multifaceted business impact, directly addressing key areas of cost, quality, and speed.

- Drastic Efficiency Gains in Maintenance and Setup: By making machine verification 4 times faster, the solution fundamentally changes the machine maintenance process. It transforms a lengthy, disruptive procedure that often requires specialized technicians into a rapid, automated routine that can be performed by the operator. This dramatically increases machine availability for production and reduces operational costs.
- Substantial Reduction in Quality Costs: An 80% reduction in production scrap has a direct and massive financial impact. It minimizes wasted materials, energy, and machine time. By catching and correcting errors in-process, the

system prevents the production of faulty parts, leading to higher first-time-right rates, improved process reliability, and enhanced customer trust.

- Increased Throughput and Competitiveness: The 10% improvement in overall production cycle time allows the company to produce more with its existing assets. This boosts manufacturing capacity, shortens lead times to customers, and increases the factory's overall agility and competitiveness in the market.

In summary, the implemented solution goes beyond a simple technical demonstration; it provides a clear roadmap to a more efficient, reliable, and cost-effective manufacturing process.

3 Pilot 4: Avio Aero

3.a Business Scenarios 1 & 2

3.a.1 Full-scale implementation

Business scenarios 1 and 2 implement an automated defect detection system for mechanical components within the aeronautical manufacturing domain, addressing specific limitations inherent in the manual inspection processes at Avio Aero.

Currently, visual inspection at Avio Aero (business case 1) is conducted entirely manually, relying on human operators' expertise without the aid of digital tools or intelligent decision-support systems. This manual approach introduces several critical challenges, including variability in defect classification due to subjective human interpretation and an increased risk of human error. To overcome these challenges, the business case focused on developing and training a machine learning (ML) model capable of analysing images of components and automatically identifying regions likely to contain defects.

In addition to its operational use for automated defect detection, the trained model has also been exploited as a training tool to support the education of new maintenance personnel (business case 2). Specifically, a training module has been developed to support practical test/practice sessions, in which trainees are asked to analyse images of components and identify potential defects. Their responses are then compared against the predictions made by the ML model, which serve as objective references. Each trainee receives a score based on the accuracy and completeness of their responses, allowing for an objective skill assessment and a learning focused on the most challenging cases. This approach helps reduce subjectivity in the learning process and contributes to standardizing operator training.

Given the distributed nature of the industrial environment - where multiple factories or production lines manage different components and maintain locally stored - often confidential datasets - the project employs a Federated Learning (FL) paradigm. FL facilitates collaborative model training without centralizing sensitive data. Each production site trains a local model on its proprietary dataset and shares only model updates (e.g., gradients or weights) with a central server. The server aggregates these updates to generate a shared global model. This decentralized approach preserves data privacy and complies with internal policies and external data protection regulations, while simultaneously enhancing the diversity and representativeness of the training data.

3.a.1.1 Architecture

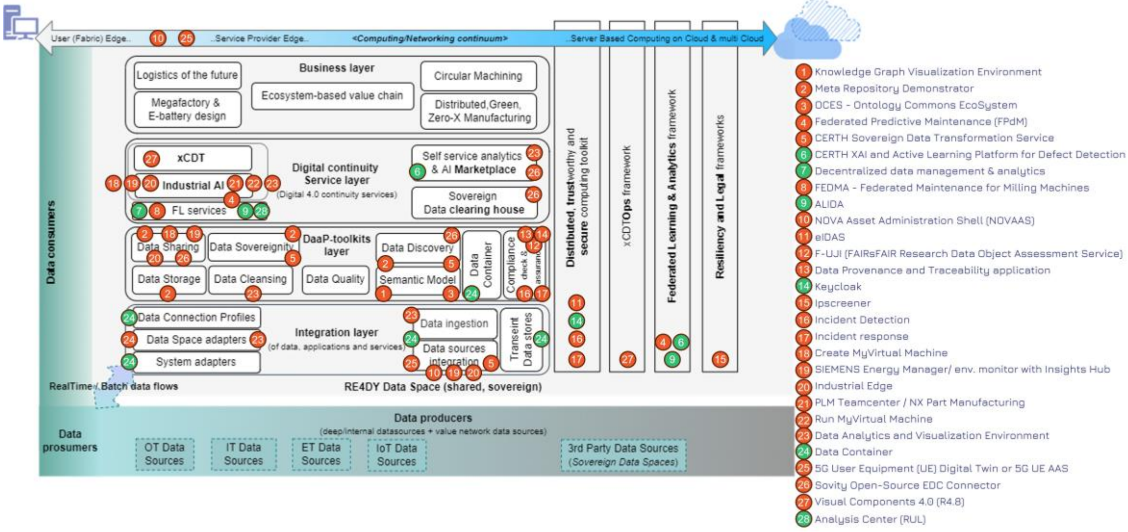


Figure 35 – RE4DY Toolkit <-> Reference Architecture Mapping

The RE4DY Toolkit components (Figure 35) identified during the previous design and implementation phases, have proved well suited for the realisation, execution and validation of the solutions for business cases 1 and 2. Specifically, for these first two scenarios the following components have been adopted and further developed:

- Component 6: CERTH XAI and Active Learning Platform for Defect Detection.
- Component 9: ALIDA.
- Component 14: KeyCloak.

The implemented solutions are the result of the mutual integration between these components that, by working together, provide a seamless end-to-end user experience to data scientists and quality inspectors alike. The relations between the components selected for the Avio Aero use cases are depicted in Figure 36 below:

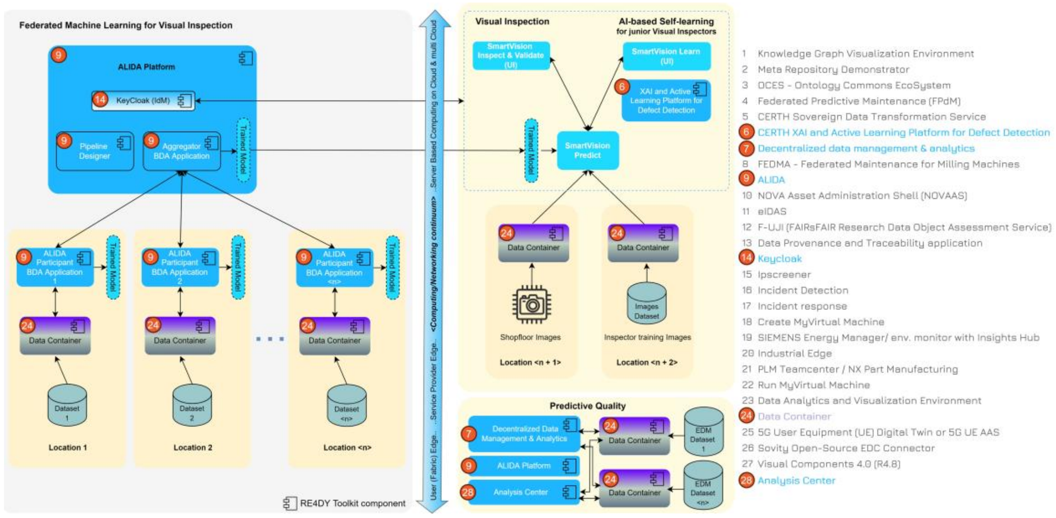


Figure 36 – Diagram showing the relations between the RE4DY toolkit components selected for the Avio Aero use case

In particular, component 6 – the XAI service – complements the AI model for defects detection by providing further insights into the behaviour of the model. It does so by returning heatmaps highlighting the level of attention that the model has given to areas of the image. Component 9 – ALIDA – has provided data scientists with a convenient way to develop the AI/ML pipelines, supporting their deployment through docker. This component has also been extended with the Smart Vision Suite: a satellite set of applications enabling quality inspectors to view and manipulate model results. Finally, access to the ALIDA platform has been securely managed by Component 14 – KeyCloak – which integrates with the project's Single Sign-on (SSO) system to provide access to ALIDA.

As for the enclosing deployment architecture (Figure 37), what follows describes how this has changed also highlighting its main strength points. More details on the specific RE4DY Toolkit components and how they have been further enhanced will be provided in the next sections.

To start with, below (Figure 37) is highlighted - in green - the portion of architecture that has changed.

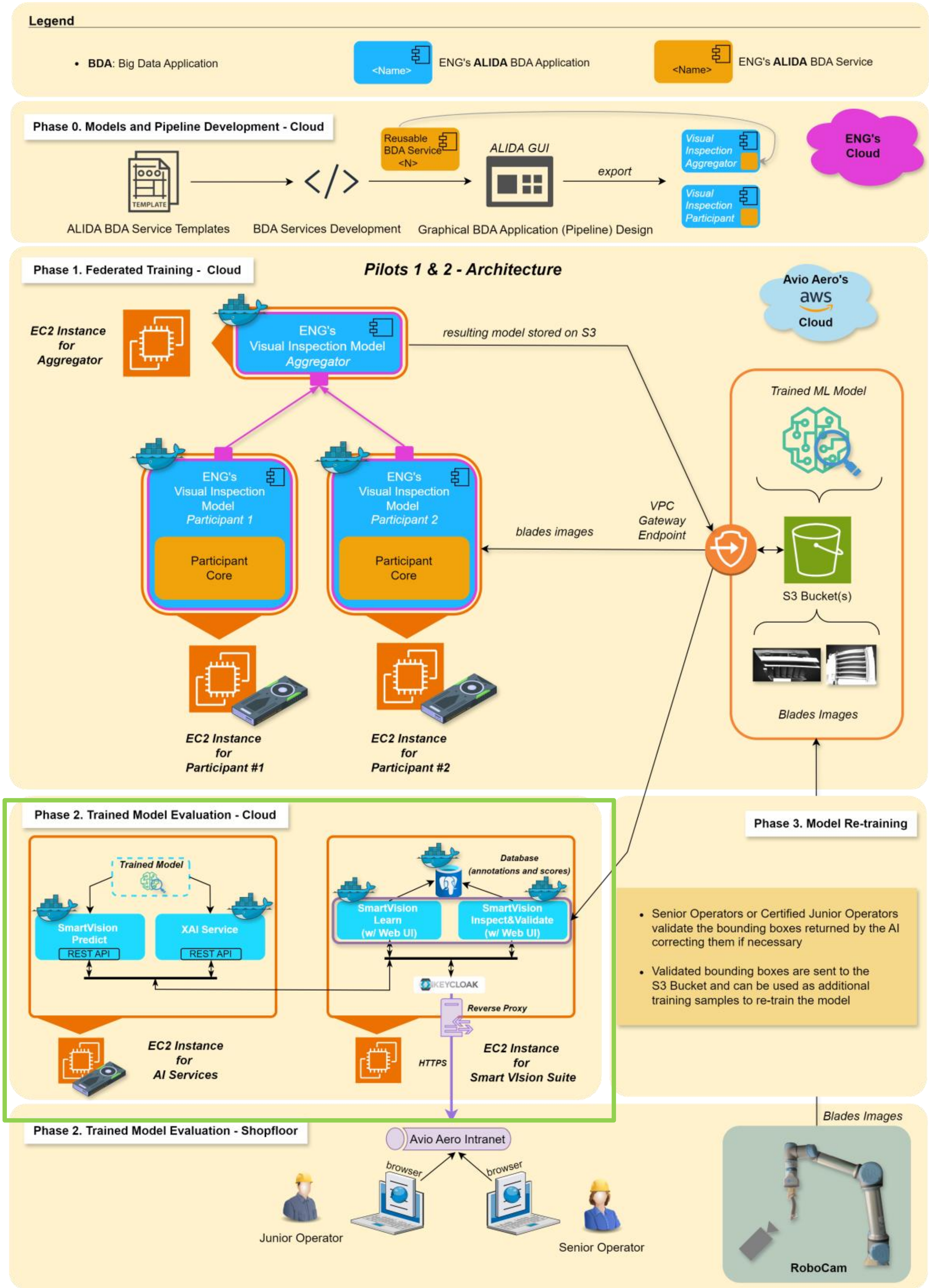


Figure 37 – Business Cases 1 & 2 Deployment Architecture

In particular, it can be noticed that the execution of the Predict and XAI services now occurs on a GPU-equipped node. These services, originally deployed on the same VM hosting the Smart Vision Suite, happened to require more computational resources, hence the need to move to a higher performing node. This possibility of rearranging workloads demonstrates the flexibility of both architecture and solutions.

As for the development of the AI solutions, ALIDA has been successfully adopted by the data scientists to develop the BDA Applications (pipelines) using existing or customly built BDA Services (pipeline blocks, Figure 38). As a result, the ALIDA catalogue has been enriched with both custom-built BDA Services and ready-to-use BDA Applications, from where they can be shared with other platform users when necessary.

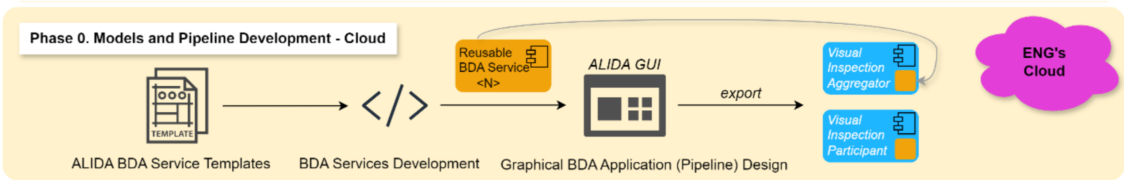


Figure 38 – ALIDA pipeline development workflow

From a computational resources' standpoint, the availability of virtual machines with high-end GPUs (Figure 39) has enormously sped up not only model training, but also the preliminary data analysis tasks and model inference. In the latter shortening the time required to obtain the predicted defects annotations during visual inspection in production or junior operator learning phase.

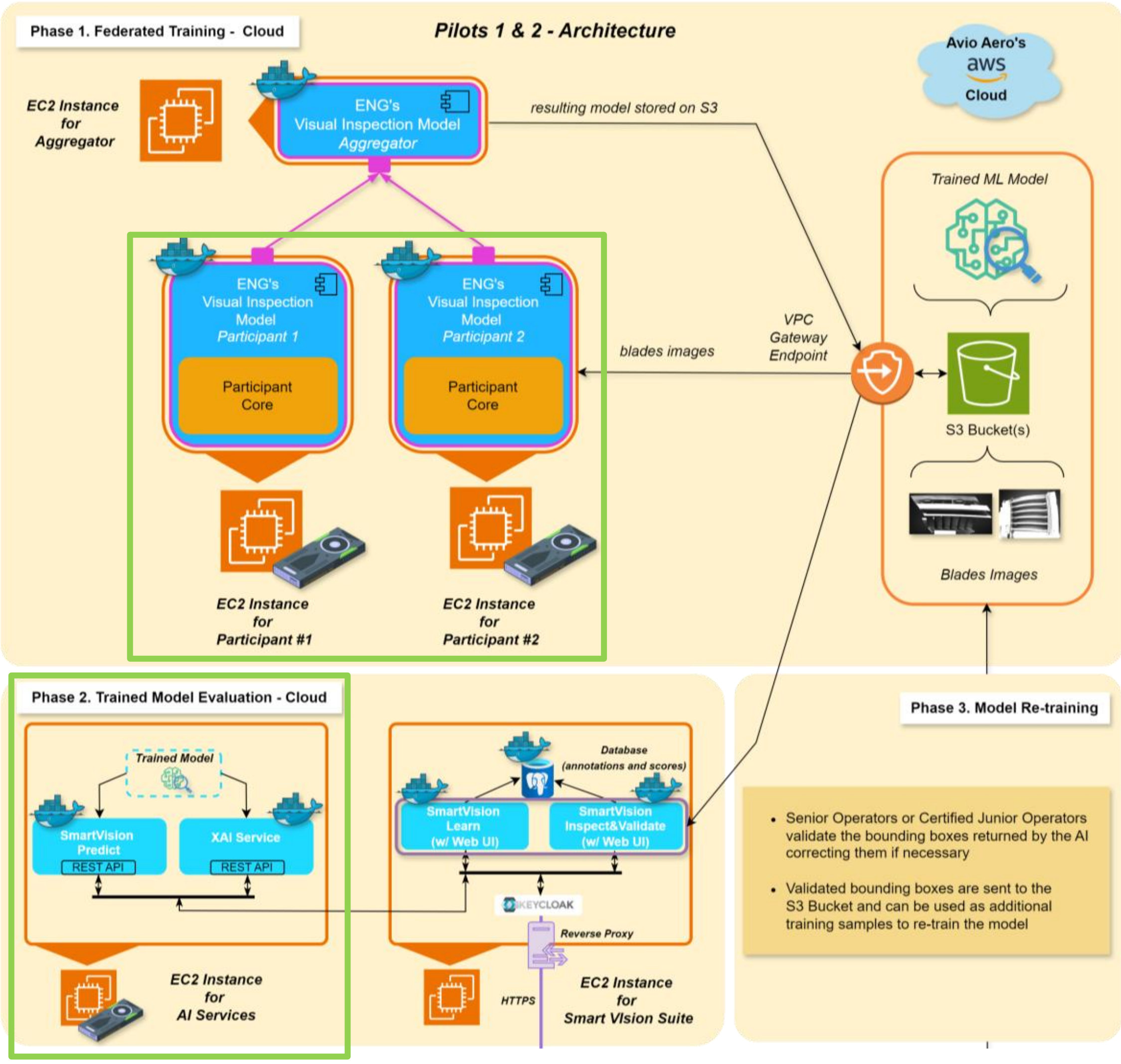


Figure 39 – GPU-equipped VMs for training and inference phases

The same diagram (Figure 39) also shows a widespread use of dockerized ALIDA BDA Applications. The use of docker has noticeably simplified and accelerated the deployment and update of the solutions, also allowing for a quick redistribution of the workload across the nodes.

Cybersecurity-wise, the arranged AWS virtual machines, network configurations and tools, have guaranteed safe access to data and computational nodes, both from the inside and outside of the Avio Aero network.

3.a.1.2 AI Models

3.a.1.2.1 AI-based Defects Detection

Dataset and defects labelling

The dataset consists of 1,488 high-resolution images, evenly divided between 744 non-defective samples and 744 images annotated with at least one defect. The distribution of defect types is as follows:

- BRAZING SPOTS: 163 images (most frequent)
- POSITIVE METAL: 139 images
- CUT BACK: 108 images
- DENT: 102 images
- CAVITY: 92 images
- DEFORMATION: 76 images
- HIGH METAL: 64 images (least frequent)

This distribution reveals a significant class imbalance, with the most frequent defect type (BRAZING SPOTS) appearing in more than twice as many images as the least frequent (HIGH METAL). The limited sample size and such class imbalance poses a major challenge for training machine learning models, as models tend to become biased toward the majority classes, leading to poor generalization and underperformance on underrepresented defect types. Based on best practices and empirical evidence in object detection, robust model training and reliable defect recognition typically require at least 1,500 images per class, along with a minimum of 10,000 instances (bounding boxes). The current dataset falls short of these benchmarks, potentially resulting in suboptimal detection accuracy.

In addition to limited data volume and class imbalance, the dataset exhibits a critical annotation-related limitation: each image is labelled with only one defect type, even when multiple defects are visibly present. This incomplete labelling introduces semantic ambiguity and prevents the model from learning to detect multiple defect types that may occur simultaneously. In safety-critical domains like aerospace, such limitations can significantly compromise the reliability of automated inspection systems.

To address this, future annotation efforts should adopt a multi-label, instance-level annotation approach, ensuring that all visible defects are accurately and consistently labelled to support more comprehensive and robust model training.

Image Acquisition and Labelling: The data collection leveraged a custom-built industrial vision system designed for seamless integration within Avio Aero's in-line production processes. The system captures high-resolution images (5472×3648 pixels) of mechanical components under controlled positional and lightning configurations. This setup minimizes variability during acquisition, ensuring consistent visual conditions, and enhancing data quality for model training. Each inspection targets a single mechanical component, uniquely identified by its serial number. For every inspection, a fixed set of images was captured from predefined perspectives, with each view explicitly designed to highlight a specific Region of Interest (ROI) on the component surface.

The images adhere to a structured naming convention including an inspection identifier and a view identifier, enabling precise traceability of views per component. Binary masks were also provided for each image to isolate the ROI, filtering out irrelevant background or structural elements.

Defect labelling was performed manually by experienced operators at Avio Aero. For each image, operators identified and annotated surface defects within the defined ROIs. Annotations were organized in a file mapping filenames to their associated defects. Each defect is represented by a defect code and a bounding box specifying the position and size of the defect within the image. An important characteristic of the annotated defects is their size. In most cases, defects occupy only a very small portion of the total image area, making them visually subtle and difficult to distinguish from the background. This is especially true for defect types such as BRAZING SPOTS, POSITIVE METAL, CAVITY, which are represented by extremely small bounding boxes.

These characteristics have been carefully considered during preprocessing and model design, as they directly affect the model ability to learn and detect small-scale anomalies.

Proposed Methodology

The preprocessing pipeline and model architecture were carefully designed to address the specific challenges identified in the dataset analysis—namely, the small size of defects, class imbalance, and incomplete annotations. These constraints necessitated a strategy that enhances defect visibility and supports scalable inference in high-resolution industrial settings.

Data processing Strategy: Given that most defects occupy only a small portion of the image, training on full-resolution images (5472×3648 pixels) would dilute the signal of interest, making it difficult for the model to learn meaningful features. Therefore, images were divided into overlapping 640×640-pixel patches, which increased the relative size of defects and improved their visibility.

To reduce noise, crops were generated only within the ROI masks. Given the dominance of non-defective areas, a selective sampling strategy was employed—retaining all crops containing defects and sampling a subset of defect-free crops. This balance mitigates class imbalance and prevents the model from being biased toward background predictions.

Data augmentation techniques such as random flips, rotations, and brightness/contrast adjustments were applied dynamically during training to improve model generalization and robustness.

Model Selection: The object detection model selected for this project is YOLOv8. YOLOv8 was selected for its proven performance in object detection tasks, particularly in detecting small and sparse objects. Its anchor-free architecture and improved feature fusion mechanisms make it well-suited for identifying small defects in high-resolution industrial images.

The model uses a convolutional backbone comprising Cross Stage Partial (CSP) modules and lightweight C2f (Concatenate to f) blocks to enhance gradient flow and reduce redundancy. A feature aggregation neck and decoupled detection head enable multi-scale feature extraction and improved convergence. A Keras-compatible implementation was chosen for its flexibility in customizing training pipelines.

Inference Method: While training is performed on localized crops, real-world deployment requires inference on full-resolution images. To bridge this gap, the system integrates Slicing Aided Hyper Inference (SAHI):

- Slicing: During inference, each high-resolution image is divided into overlapping patches of the same dimensions used during training (640×640 pixels).
- Model Prediction: The trained model is applied independently to each image slice.
- Prediction Merging: Predictions from all slices are merged to produce a unified detection output.

This approach ensures that small and spatially sparse defects are not missed, while maintaining the scalability and efficiency required for industrial inspection workflows.

Pipelines

Preprocessing Pipelines: The following section details the steps taken to prepare the data for training the model:

1. ROI Masking: Binary masks provided by Avio Aero define ROIs. Pixels outside these regions are zeroed out to remove background noise. Bounding boxes outside the ROIs are discarded to maintain label consistency.
2. Crop Generation: A sliding window with 320-pixel overlap generates 640×640 crops. For defective images, only crops containing bounding boxes are retained. For non-defective samples, all ROI-aligned crops are kept. This ensures sufficient positive and negative samples.
3. Dataset Split: The dataset is divided into 80% training, 10% validation, and 10% testing. Crops from the same original image are kept in the same split to prevent data leakage. Class distributions are balanced across subsets.

Model Training Pipeline: The following section presents the steps taken to prepare pre-processed data for training the model, the training strategies, and the model evaluation.

1. Crop Sampling: To address the high prevalence of background-only crops, only a subset of non-defective crops is used. All crops with annotated defects are retained to maximize learning from scarce positive samples.
2. Training: YOLOv8 is trained using sampled crops with on-the-fly augmentations. Early stopping based on validation metrics prevents overfitting and ensures optimal performance.
3. Threshold Tuning: The model output is filtered using two key parameters: the confidence threshold and the Intersection-over-Union (IoU) threshold. The confidence threshold sets a minimum score required for the model to consider a prediction as valid, helping to eliminate uncertain or spurious detections. The IoU threshold is used during Non-Maximum Suppression (NMS) to determine whether

predicted overlapping bounding boxes should be merged, which prevents multiple detections of the same defect. Both thresholds are fine-tuned on the validation set to maximize the F1-score, ensuring a balanced trade-off between precision and recall.

4. Evaluation: Model performance is assessed using the SAHI framework, which enables robust inference on high-resolution test images. To determine detection accuracy, standard object detection metrics are reported: precision (the proportion of predicted defects that are correct), recall (the proportion of actual defects that are detected), and F1-score (the harmonic mean of precision and recall). A predicted bounding box is considered a true positive if the IoU with the ground truth bounding box is at least 25%.

Experiments

To validate the proposed defect detection approach and understand its behaviour under different levels of complexity, two experiments were conducted: Single-Defect Detection and All-Defects Detection. The first experiment simplified the task by focusing on a single, visually distinct defect type, while the second introduced the full range of available annotations to simulate more realistic conditions. This progression – from simple to complex – helps assessing the model performance in both controlled and practical scenarios.

Single-Defect Detection

The first experiment focused solely on detecting the cut back defect. This type was chosen because it is typically larger, more visually distinct, and easier to identify than other surface anomalies in the dataset. For this experiment, only the bounding boxes corresponding to cut back defects were retained in the training set; all other annotations were excluded. This served as a controlled baseline to evaluate the model ability to detect a well-defined defect under simplified conditions.

Data: To ensure the model was trained effectively on the target defect, the following sampling strategy was applied:

1. Positive Samples (Cut Back Defects): All crops containing at least one annotated cut back defect were included in the training set. This ensured the model had full exposure to every available example of the target class.
2. Other-Defect Samples (Treated as Background): A number of crops equal to 50% of the positive samples were randomly selected from images containing other types of defects. These other defects were not annotated, so the model treated them as background.
3. Background-Only Samples: Another 50% (relative to the cut back samples) were randomly selected from crops with no annotated defects at all. These served as pure background examples to help the model distinguish defect-free regions.

This sampling strategy resulted in a dataset composed of 1,044 training crops, 48 validation crops, and 124 test crops, all tailored to evaluate the ability of the model to detect a single, well-defined defect under controlled conditions.

Results: Bounding box level metrics evaluate the model ability to detect and localize each individual defect. Each predicted bounding box is compared against ground truth annotations to determine whether it correctly identifies a defect. This provides a detailed view of how well the model performs at the object level, especially when multiple defects appear in a single image.

From the performance metrics in Table 5, we see that:

- Precision (0.26): Only 26% of the predicted bounding boxes corresponded to actual defects. The model frequently misclassified background regions or other defect types as Cut Back, resulting in a high number of false positives.
- Recall (1.00): The model successfully detected all actual Cut Back defects. This is especially important in safety-critical contexts like aerospace, where missing a defect is unacceptable.
- F1-score (0.41): This metric balances precision and recall. While the model demonstrates excellent sensitivity, its low precision reduces overall reliability.

Table 5 – Single-Defect Detection MODEL Performance Metrics

Metric	Value
PRECISION	0.26
RECALL	1.00
F1-SCORE	0.41

The confusion matrix in Figure 40 reveals that:

- The model correctly identified 11 Cut Back defects but also misclassified 2 Deformation and 1 High Metal defects as Cut back.
- Background regions were incorrectly predicted as Cut Back, significantly contributing to the false positive count.

Bounding Boxes Confusion Matrix

Predictions \ Ground Truths	BRAZING SPOTS	CAVITY	CUT BACK	DEFORMATION	DENT	HIGH METAL	POSITIVE METAL	BACKGROUND
CUT BACK	0	0	11	2	0	1	0	29
BACKGROUND	24	5	0	6	11	7	17	-

Figure 40 – Single-defect detection mode confusion matrix

Examples: To illustrate the performance and behaviour of the trained model, representative prediction examples are presented in this section. In the figures below:

- Green bounding boxes represent the ground truth annotations.
- Blue bounding boxes represent the model's predictions.

These examples were selected from the test set and demonstrate the model's ability and limitations to detect cut back defects.

Correct Prediction – Cut Back Defect Detected

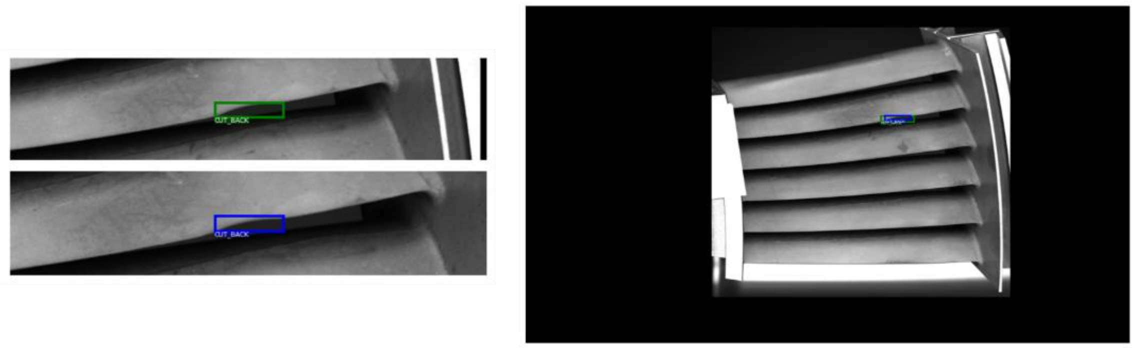


Figure 41 – Single-defect Model – Correct Prediction

Figure 41 shows an example of a correct prediction, where the model properly identified a cut back defect. The predicted bounding box closely matched the ground truth annotation.

Wrong Prediction – Background Misclassified as Defect

Figure 42 – Single-defect model – Background misclassified as defect

As shown in Figure 42, the model incorrectly predicted a cut back in a region with no annotated defect. This false positive likely stems from background textures that visually resemble known defect patterns, indicating a need for more diverse background examples during training.

Wrong Prediction – Wrong Defect Type Detected

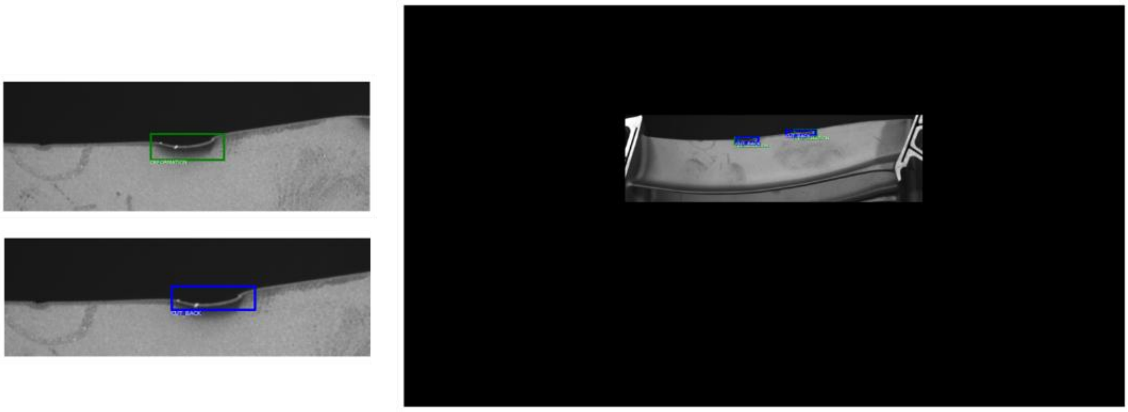


Figure 43 – Single-defect model – Wrong defect type

In Figure 43, two deformation defects were misclassified as cut back defects. Although the model correctly localized the anomalies, it assigned the wrong class label. This suggests difficulty in distinguishing between visually similar defect types, especially when training data is limited.

Labelling Ambiguity – Prediction Without Ground Truth

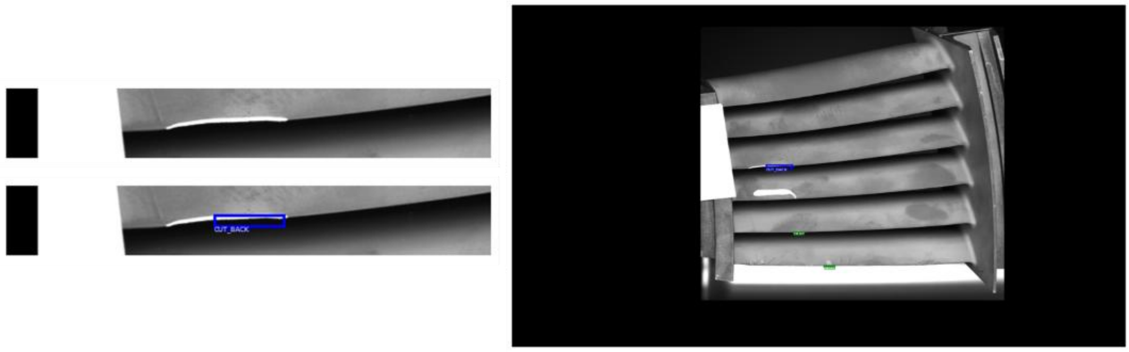


Figure 44 – Single-defect model – Prediction without ground truth

Figure 44 illustrates a case where the model predicted a Cut Back defect in a region that appears visually defective. The image contains two annotated Dent defects, but no annotation for the predicted Cut Back. Upon inspection, the prediction appears valid, suggesting the presence of a third, unlabelled defect within the Region of Interest (ROI). This example highlights a known limitation in the dataset: only one defect type was annotated per image, even when multiple types were visibly present. As a result, valid predictions like this are incorrectly counted as false positives.

Summary: This experiment demonstrates that the model is highly sensitive to detecting Cut Back defects, achieving perfect recall with no missed detections. However, it also shows limited specificity, frequently misclassifying background regions or other defect types such as Cut Back, resulting in a high number of false positives.

While this trade-off may be acceptable in early-stage or safety-critical applications - where false positives are preferable to false negatives - it highlights critical areas for improvement:

- Annotation Quality: Some false positives may be attributed to missing or incomplete labels in the dataset. In some cases, the model correctly identifies defects that were not annotated, which are then incorrectly counted as false positives. This underscores the need for a more comprehensive and consistent annotation strategy that captures all visible defects in each image.
- Background Sampling Strategy: The current training setup includes only a sample of background-only crops. Further experiments should be conducted to fine-tune the ratio of background to defect-containing samples in the training set. Increasing the number of background crops or improving their selection could help reduce the false positive rate without compromising specificity.
- Limited Dataset Size: The overall volume of annotated data is relatively small, especially when considering the diversity of defect types and the need for robust generalization. The limited number of examples per class restricts the model's ability to learn nuanced distinctions between defects and background. Expanding

the dataset—both in terms of the number of images and the variety of annotated defects—is essential for improving model performance and reliability.

All-Defects Detection

In the second experiment, the model was trained using all available annotations in the dataset. Unlike the first experiment, no filtering was applied—every labelled defect, regardless of class, was included in training. This configuration represents the target use case for the system: detecting diverse surface anomalies within high-resolution industrial images.

Data: The following sampling strategy was applied:

1. Positive Samples: All crops containing at least one annotated defect.
2. Background Samples: Another 50% (relative to the positive samples) were randomly selected from crops with no annotated defects.

This sampling strategy resulted in a dataset composed of 4,213 training crops, 492 validation crops, and 571 test crops.

Results

Table 6 – All-DEFECTs DETECTION MODEL PERFORMANCE METRICS

Metric	Value
PRECISION	0.20
RECALL	0.36
F1-SCORE	0.26

From the performance metrics in Table 6, we see that:

- Precision (0.20): Only 20% of the predicted bounding boxes were correct. The model frequently misclassified background or other visual patterns as defects, leading to a high number of false positives.
- Recall (0.36): The model detected just over one-third of the actual defects. While this is a drop from the single-defect experiment, it reflects the increased complexity of detecting multiple defect types.
- F1-score (0.26): This score reflects the trade-off between low precision and moderate recall, indicating that the model struggles to balance sensitivity and specificity in a multi-class setting.

The confusion matrix in Figure 45 reveals that the model often misclassifies defects or fails to detect them altogether. While some correct predictions are made, a significant number of false positives and false negatives are observed across all defect categories. This indicates that the model has difficulty distinguishing between different defect types and background regions, especially when defects are small, visually similar, or underrepresented in the training data.

Bounding Boxes Confusion Matrix

Predictions \ Ground Truths	BRAZING SPOTS	CAVITY	CUT BACK	DEFORMATION	DENT	HIGH METAL	POSITIVE METAL	BACKGROUND
BRAZING SPOTS	10	0	0	0	0	0	3	47
CAVITY	0	0	0	0	0	0	0	6
CUT BACK	0	0	11	1	0	1	0	13
DEFORMATION	0	0	0	3	2	0	0	25
DENT	0	0	0	0	2	0	0	8
HIGH METAL	0	0	0	0	0	1	0	2
POSITIVE METAL	2	2	0	0	0	0	3	15
BACKGROUND	12	3	0	4	7	6	11	-

Figure 45 – All-defect detection model confusion matrix

Examples: This section presents representative prediction examples to illustrate model's behaviour when trained on all available defect types. The same visual convention is used:

- Green bounding boxes indicate ground truth annotations.
- Blue bounding boxes indicate model predictions.

Correct Prediction – Small Defects Detected

Figure 46 shows the successful detection of three small Positive Metal defects. Despite the very limited pixel area occupied by the defects, the model accurately localized the three defects. The bounding boxes closely align with the ground truth annotations, demonstrating the model's ability to detect subtle surface anomalies in high-resolution images.

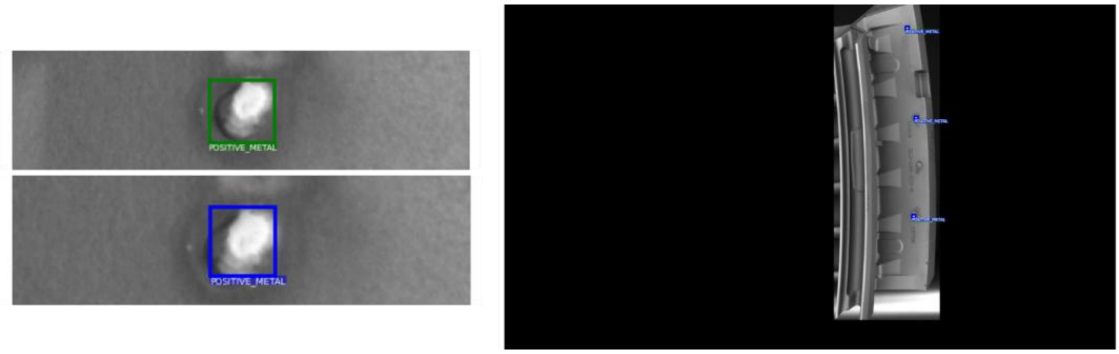


Figure 46 – All-defect model – Correct Predictions

Labelling Ambiguity – Small Defects Not Annotated

Figure 47 presents a case where the model predicted two small defects: one labelled as Brazing Spots and another as Positive Metal. However, the image contains no annotated defects. While it is not possible to confirm the correctness of these predictions without expert validation, the predicted regions appear visually consistent with known defect patterns. This suggests a likely case of missing annotations, where valid predictions are incorrectly counted as false positives.

Figure 47 – All-defect detection model – Predictions with no ground truths

Summary: This experiment highlights the challenges of scaling from single defect to multi-defects detection. While the model retains some ability to detect defects, its performance drops significantly in terms of both precision and recall. The high number of false positives suggests that the model is overly sensitive to visual patterns that resemble defects, while the high number of false negatives indicates that many actual defects are being missed.

The issues identified in the single-defect experiment remain relevant here—particularly the need for more comprehensive annotations and a better-tuned background sampling strategy. However, the multi-defect setting introduces additional complexity due to the low number of examples per class, especially across the wide variety of defect types. This scarcity limits the model's ability to generalize and accurately distinguish between different defect categories and background features.

Despite these limitations, the approach remains promising. The model demonstrates the ability to detect even very small and subtle defects, which is a critical capability in high-precision manufacturing environments. With further improvements in data quality, class balance, and training strategies, the system has strong potential to evolve into a robust and scalable solution for real-world industrial inspection.

The study described has demonstrated that automated defect detection using deep learning holds strong potential for aeronautical manufacturing, particularly in identifying even very small and subtle defects that are often missed during manual inspection. However, the current system still faces limitations that must be addressed to improve its reliability.

One of the most pressing issues is the quality and completeness of the annotations. In many cases, only a single type of defect is labelled per image, even when multiple defects are visibly present. This not only limits the model's ability to learn from co-occurring defects but also leads to false positives when the model correctly identifies unlabelled defects. A more comprehensive and consistent annotation strategy is needed to ensure all visible defects are accurately captured.

Another significant constraint is the limited size of the dataset. The number of images and defect instances per class falls short of the thresholds typically required for robust object detection. This restricts the model's ability to generalize and to distinguish between fine-grained defect variations and background patterns. Expanding the dataset in both volume and diversity is essential to improve model robustness.

Further parameter tuning and experimental validation are necessary. Optimizing model thresholds and refining data sampling strategies will be key to achieving a better balance between sensitivity and precision. The current training configuration includes a sample of background-only crops. Further experiments are needed to fine-tune the ratio of background to defect-containing samples. Increasing the number of background crops or improving their selection could help reduce the false positive rate without compromising specificity.

With the federated learning infrastructure already in place, future work can also explore scaling the system across multiple production sites. This would enable collaborative model training while preserving data privacy, ultimately supporting a more scalable and secure deployment of automated defect detection in industrial environments.

Explainable AI (XAI)

To complement the AI model for defects detection by providing further insight into the behaviour of the model, heatmaps are returned that highlight the level of attention given by the model to areas of the image. This is achieved by further optimizing the D-RISE XAI technique[3].

[3] Petsiuk, V., et al.: Black-box explanation of object detectors via saliency maps. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 11443–11452 (June 2021)

The heatmap is intended to be read as a visual explanation of the decision. Stronger influence is indicated by warmer colours such as red and yellow, while weaker influence is indicated by cooler colours such as blue. A bright region indicates where the model found the most evidence for a defect. The heatmap does not by itself confirm that a defect exists. Rather, it shows where the model looked in order to reach its decision.

In practice, good heatmap alignment is expected to overlap with the annotated defect region in the ground truth. When attention appears in unrelated areas, potential issues may be indicated, such as confusing background patterns, gaps in the annotations, or model shortcuts. These signals are used to guide data review, support model debugging, and communicate results to engineers and domain experts in an accessible way.

In details, the D-RISE explainable AI technique has been extended to support models stored in h5 format, thereby broadening its applicability across a wider range of neural network frameworks. Soft masks are now generated through interpolation rather than using binary masks, which allows for more nuanced attribution of pixel importance. In order to focus on the most informative regions, predicted masks whose intersection over union with the input exceeds a specified threshold are retained for further analysis. Saliency maps are accumulated by performing element-wise multiplication of each mask with its associated score before summation, replacing the previous tensor-dot approach to improve computational efficiency.

A fallback mechanism has been implemented to enable batch processing on the CPU when GPU resources are insufficient, or memory constraints are encountered. This ensures that the technique remains robust under varying hardware conditions and can be deployed in environments with limited computational power. The extended D-RISE method has been applied to a newly curated defect localization dataset, and the resulting heatmaps have demonstrated strong alignment with the ground-truth annotations, indicating that the saliency outputs accurately highlight defect regions as, it can be seen on Figure 48.

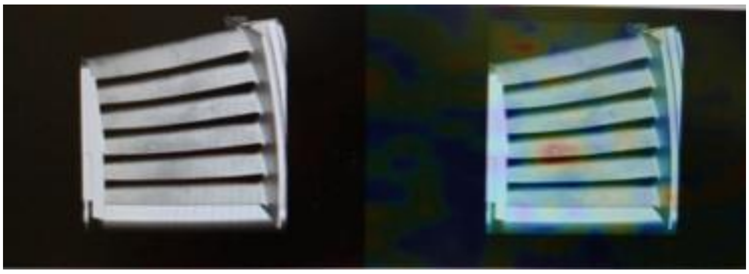


Figure 48 – Saliency Map generated by XAI indicating the RoI of the defect detection model.

An API for the extended explainable AI technique has been developed using the FastAPI[4] Python library, and an initial integration into the company's existing defect inspection pipeline has been achieved. Through this API, end users can submit images or model references and receive corresponding saliency heatmaps in a standardized format.

[4] https://github.com/fastapi/fastapi

Further validation and user-experience testing are underway to ensure seamless adoption and to refine performance under real-world conditions.

3.a.1.3 Applications

The updated Smart Vision Suite integrates with the newly introduced eXplainable AI (XAI) Service and introduces a number of enhancements to the existing features.

XAI Service

Both Inspect & Validate and Learn applications now also query the XAI Service through a new dedicated API shown below in Figure 49.

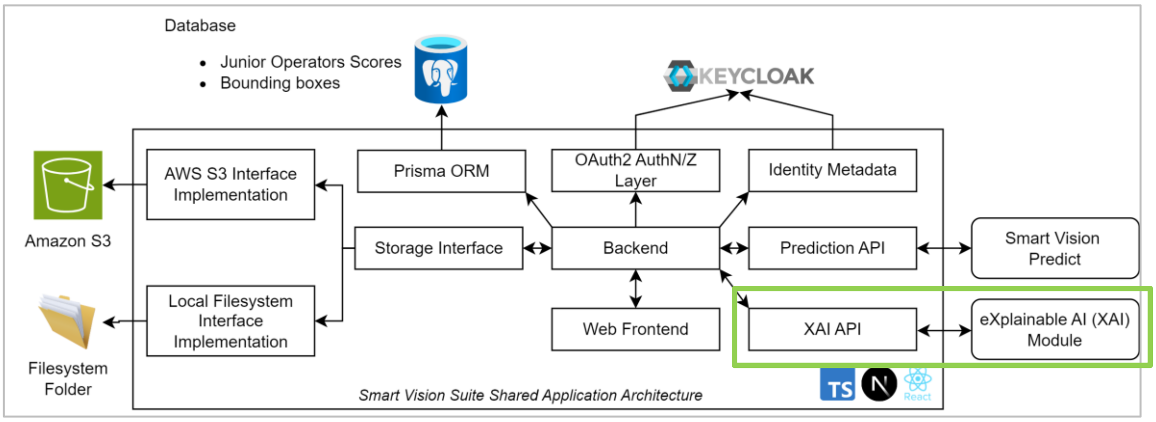


Figure 49 – Updated Smart Vision Suite architecture showing the XAI module and corresponding API

The image of the piece under inspection is sent to the service and the resulting heatmap is overlaid to the original piece image. At that point, the users can visualize the heatmap, the annotations and correlate the two sets of information to obtain a better understanding of the model behaviour; this while retaining the ability to modify the annotations as needed.

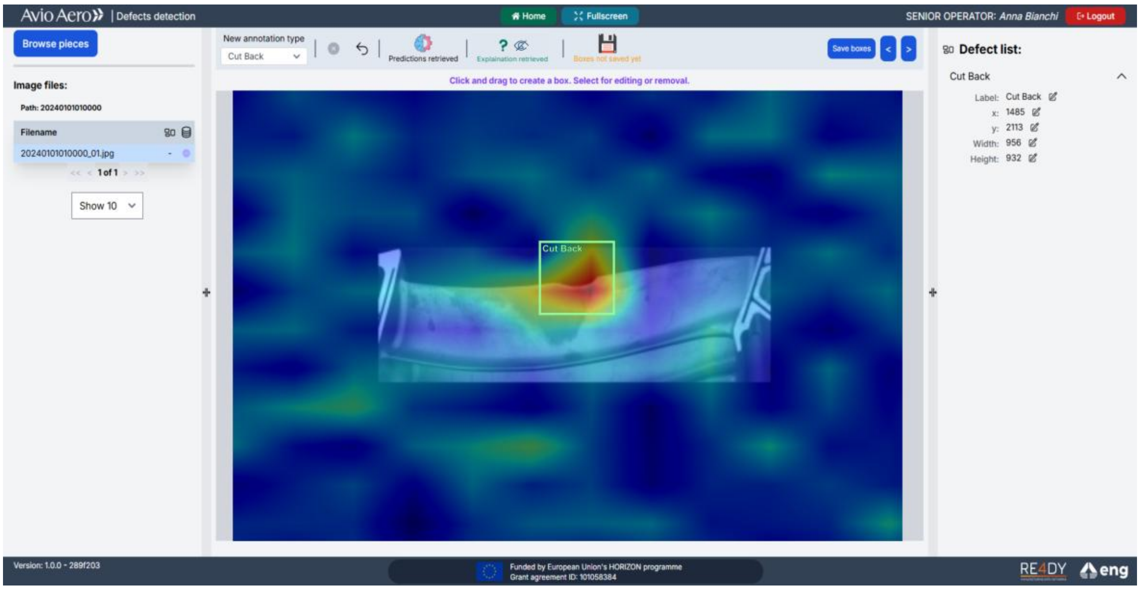


Figure 50 – Example of heatmap returned by the XAI Service superimposed to the image of the piece

On the UI/UX side, the annotator view now features a new graphical element - the *XAI Service Status Indicator* – composed of a status indicator and a button (eye icon of Figure 51):

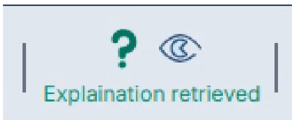


Figure 51 – XAI Service status indicator zoom in

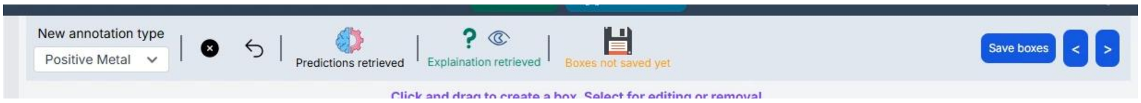


Figure 52 – Updated toolbar featuring the XAI Service status indicator

The eye button allows users to toggle the visibility of the heatmap, while the status indicator (question mark symbol and descriptive string in Figure 52) informs users about the status of AI explanation retrieval by assuming the possible statuses of:

- Retrieved: the application has obtained the explanation from the XAI service, and it is ready to show it to the user.
- Retrieving: the request has been sent, and the application is waiting for a response from the XAI service.
- Not Available: it has not been possible to obtain the explanation from the XAI service.

Updates to the Existing Features

As for the existing features, they have been enhanced to increase the robustness and maintainability of the suite.

Backend: First of all, the configuration subsystem. In the previous version, this subsystem – according the official NextJS approach – would force to rebuild the applications every time a client side-related setting had to be changed. This meant having to go through lengthy build stages and prevented changes to the configuration without access to the source code. To overcome these issues, a simple yet effective server-side loading mechanism now takes care of loading these client-specific settings from the configuration file and pass them to the browser as needed. The result is a more effective, flexible, faster and cleaner configuration experience. Another backend-side update concerned the logging sub-system. Its broader adoption across the code base, enables a closer traceability of the operations, especially useful in case of troubleshooting or auditing. Regarding the interaction with the data stores, the AWS S3 storage interface implementation has been equipped with an automatic access token refresh mechanism. Finally, changes to the docker deployment configuration allow for a better network segregation of backend services, thus increasing the level of security.

CI/CD: A GitLab CI/CD pipeline ensures that every time a push is made, a new docker image for the suite is built. This way, when the suite needs to be updated, it is generally sufficient to adapt the configuration file, pull the new image and restart the docker service to have the latest version up and running. Moreover, a set of automatic end-to-end tests developed using the Playwright framework relieves the user from manually testing the applications speeding up the development process. Moreover, these tests can also be run as part of a CI/CD pipeline, where a dedicated job executes them remotely against an instance of the Smart Vision Suite.

Deployment: A distribution package – composed of a predefined tree of directories, configuration files, docker compose files and scripts – has been developed. This package simplifies the deployment on new hosts ensuring the reproducibility of the process. Some of the scripts also contain utility commands that ease the prerequisite packages installation.

User Interface: The changes to the user interface make it:

- More responsive
- More efficient in the use of visual space
- Support full screen
- Feature tooltips to better guide users
- Present a more homogeneous look and feel

3.a.1.4 Key challenges and solutions for full-scale implementation

One of the main challenges was to access piece images available only at shopfloor level. Due to the strict cybersecurity regulations in place, it was not possible to obtain direct remote access to the shopfloor. To overcome this barrier, it was decided to leverage a relay S3 bucket accessible only through VPN-equipped Avio Aero laptops. The bucket, still directly connected to the shopfloor via a cybersecurity-compliant workflow, could now be well protected through the configuration mechanisms provided by AWS. Alternatively, and in case of small data extractions, the Avio Aero-developed DexTool was used to transfer data to the outside in a secure and compliant manner.

The second challenge concerned the quality of the dataset. The analysis reported in previous sections has in fact revealed these main addressable points of attention:

- Limited size of dataset
- Small size of defects
- Class imbalance

To address the limited size of the dataset, data augmentation techniques were employed. Small-sized defects were dealt with by dividing the original (high-res) images into smaller crops so as to increase the relative size of the defects. Moreover, the YOLOv8 model was detected as well-suited for the identification of small defects.

To reduce noise, crops were generated only within the ROI masks. Given the dominance of non-defective areas, a selective sampling strategy was employed—retaining all crops containing defects and sampling a subset of defect-free crops. This balance mitigates class imbalance and prevents the model from being biased toward background predictions.

3.a.2 Industrial trials of the pilot

3.a.2.1 Testing procedure and Barriers

The main barrier encountered while deploying and testing the tools on the Avio Aero IT infrastructure was the need to be compliant with the strict cyber security regulations. Access to the virtual machines and data storage could only occur from within the Avio Aero network and with Avio Aero-compliant computers. Therefore, it was necessary to assign and ship to each developer or system administrator requiring access to the infrastructure, a certified laptop bound to personal SSO credentials. From the laptop at that point was possible to access the Avio Aero network through a VPN and to the specific virtual machines through CyberArk. The use of IDEs and graphical tools were activated through a particular CyberArk configuration.

3.a.3 Final KPIs monitoring and validation

3.a.3.1 Industrial Outcomes and Lessons Learned

The Industrial Pilot implementation for Business Scenario 1, assumes different business key factors:

- Quality Improvement: More accurate and timely identification of defects, reducing the risk of human error and improving the overall quality of the produced parts.
- Operational Efficiency: Reduction in inspection times thanks to the automation of the defect detection process.
- Cost Reduction: Minimization of waste and costs associated with reworking defective parts.
- Continuous Learning: AI models can be continuously trained and improved to adapt to new types of defects or materials.

- Decision Support: Generation of useful data for analysis and strategic decisions, such as improving production processes.

Based on the performance levels achieved by the Yolo8 model during the training phase, a key outcome is the need to improve image quality during the data collection phase. Despite the study on the focus areas of geometry and the definition of the corresponding image acquisition points for the same part to be inspected, the image quality does not always allow for the correct identification of defect presence. This results in the AI models failing to uniquely recognize defect characteristics, thereby introducing a series of false positives in the model's inferences.

An in-depth analysis was conducted through a Design of Experiment (DoE), to identify the parameters impacting image quality and their optimal combination for each defect type.

Specifically, a study was completed by collecting images of turbine blades using photometric stereo technology to assess whether this technology could enhance the process of detecting small defects compared to image acquisition based on different positions and angles of the inspected part (see Figure 54).

The Dome implemented (Figure 53) has dozens of programmable LEDs which can be run independently of each other allowing to illuminate only the area specified by the user. When the photos are done with the help of dedicated software they are assembled in single image. There are few types of images specified by direction and type of information:

- Inclination Horizontal
- Inclination Vertical
- Roughness Horizontal
- Roughness Vertical

Figure 53 – Design of Experiment – Set-up

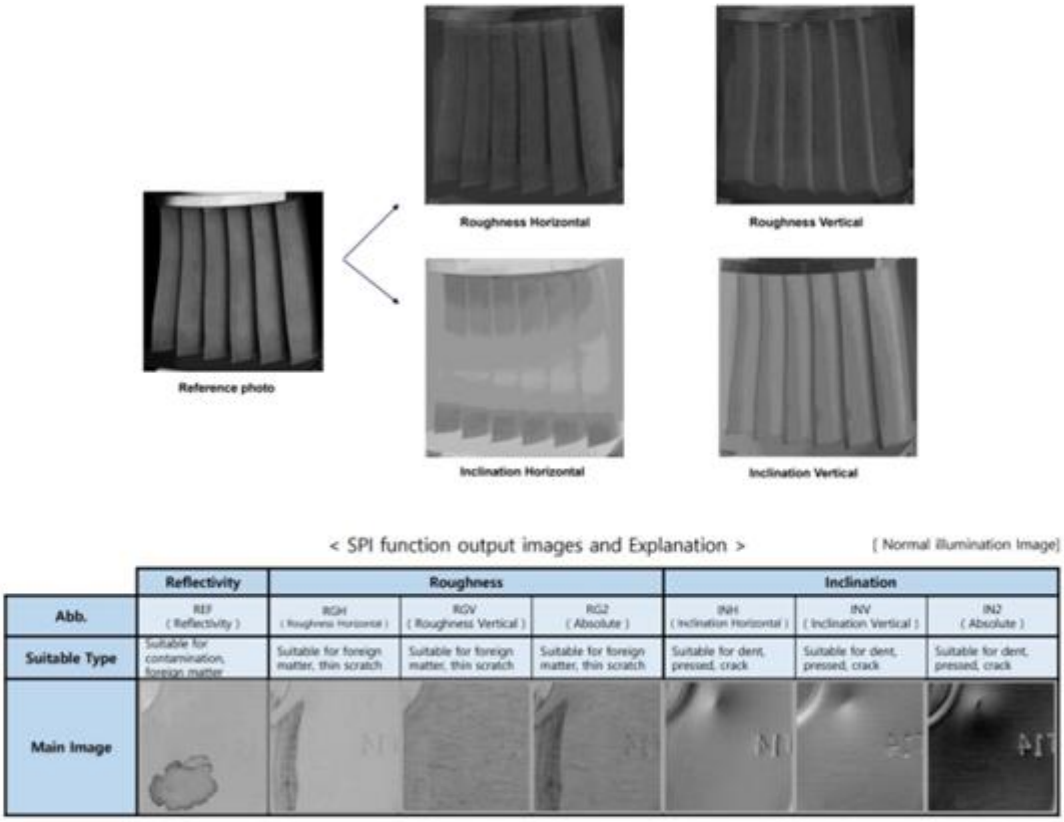


Figure 54 – Example of filters application

The Variables identified for Design of Experiment included:

- Filter Type
- Type of exposure:
- Sequence (Illuminating the area and taking photos)
- Overlay method (Camera takes data, illuminating the area, photo taken)
- Illumination correction
- On/Off (Additional illumination correction in the photo at the price of visibility of the examined area)

Figure 55 summarises the variables combined for the Design of Experiment.

Variable / Level	1	2	3	4	5	6
Filter Type	RGH	RGV	RG2way	INH	INV	IN2way
Capture method	sequence	overlap grab				
Inclination correction	ON	OFF				

Figure 55 – Summary of the variables for DoE

Several tests on various acquisition points (views) and for different defects enabled the identification of the optimal combination of parameters to highlight the characteristics of the detected defect.

In order to illustrate the approach taken, the example below outlines the process conducted for a specific view, which focuses on the central area of the blade, as shown in Figure 56:

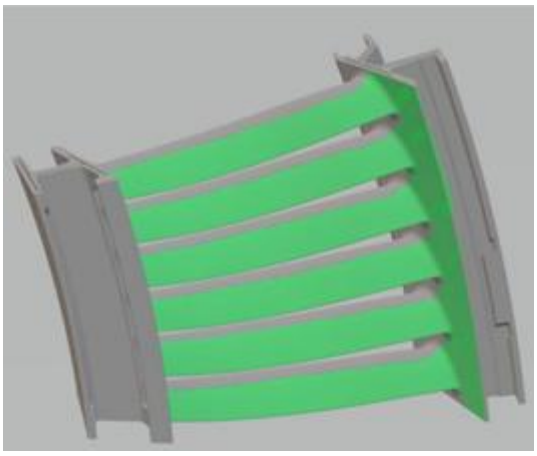


Figure 56 – Example of view

In particular, the previous view included three defects (2 positive material e 1 cut back, highlighted in Figure 57).

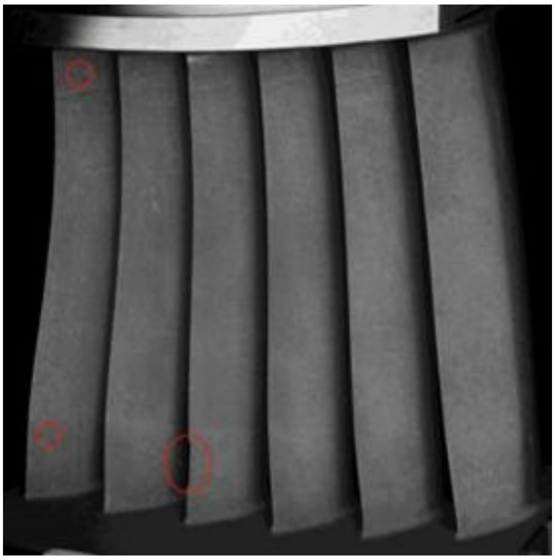


Figure 57 – Example of View with highlighted defects

In Figure 58 is reported the same image to which the optimal combination of filters has been applied, allowing for a comparison with the previously captured real image:

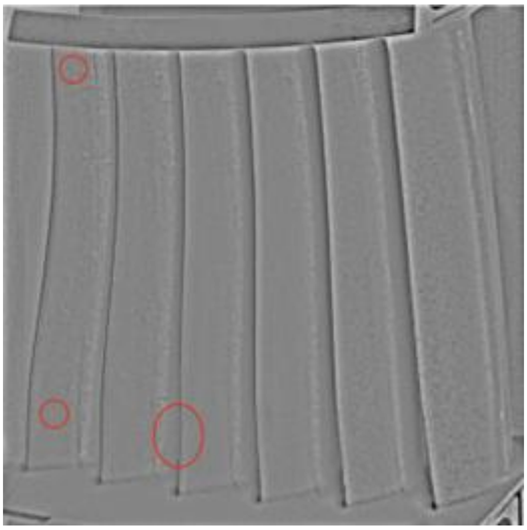


Figure 58 – Example of view with filter applied

In particular, starting from the initial image several filters have been applied, as shown in Figure 59 and Figure 60:

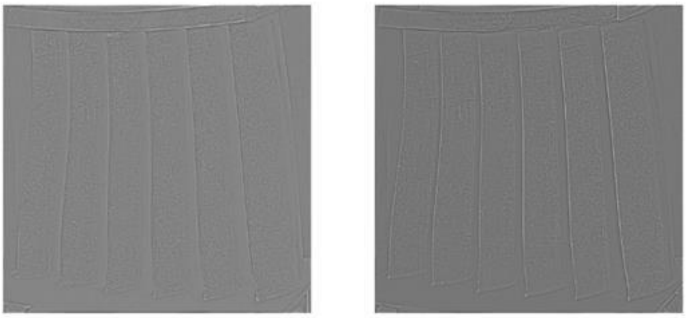


Figure 59 – Inclination Horizontal/Vertical

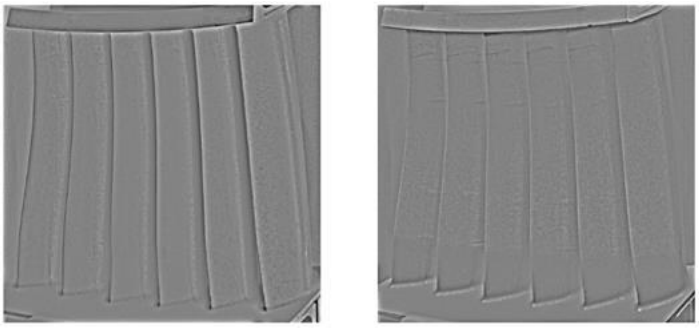


Figure 60 – Roughness Horizontal/Vertical

	REF	RGH	RGV	INH	INV

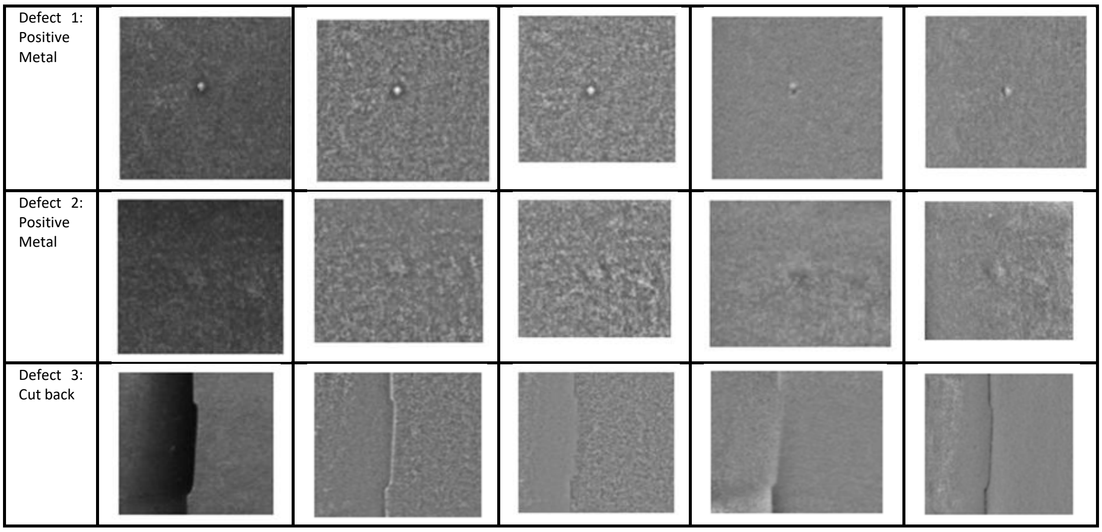


Figure 61 shows the filters comparison for each kind of defect included in the main picture:

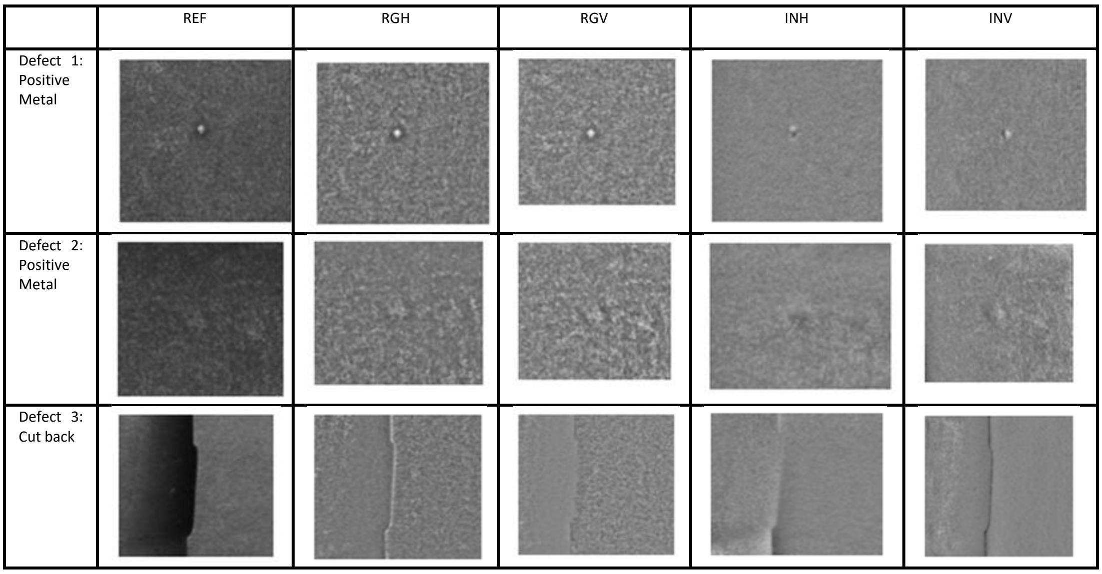


Figure 61 – Filters comparison

Following this approach for a subset of defects, it was possible to identify the optimal parameters to enhance image quality for these defects during the acquisition phase. Figure 62 reports the table of the key results:

	Type of output Images				
Type of Defect :	REF	RGH	RGV	INH	INV
Dent	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable
Cavity Pit	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable
Imperssion	^ - Weak Detection	x - Undetectable	x - Undetectable	* - Detectable	* - Detectable
Score	^ - Weak Detection	x - Undetectable	x - Undetectable	* - Detectable	* - Detectable
Scratch	^ - Weak Detection	* - Detectable	* - Detectable	^ - Weak Detection	^ - Weak Detection
Parting Line	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable
High Metal	^ - Weak Detection	x - Undetectable	x - Undetectable	* - Detectable	* - Detectable
Positive Metal	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable
Cold shut	^ - Weak Detection	* - Detectable	* - Detectable	* - Detectable	* - Detectable
Misrun	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable
Cut back	* - Detectable	* - Detectable	* - Detectable	* - Detectable	* - Detectable
Brazing spots	^ - Weak Detection	* - Detectable	* - Detectable	* - Detectable	* - Detectable
Deformation	^ - Weak Detection	^ - Weak Detection	^ - Weak Detection	* - Detectable	* - Detectable

Figure 62 – Summary table of results

This will enable more accurate results during the training phase of the models, better highlighting the characteristics of individual defects and making more effective use of the available dataset.

The Industrial Pilot implementation for Business Scenario 2 (Training Quality Inspector) revises the current certification process for junior operators in the role of inspector by proposing a training approach based on models trained to recognize defects in Pilot 1.

The advantages of having a training process for the certification of visual inspection operators based on artificial intelligence (AI) models are:

- Improved Accuracy and Quality: AI can identify defects or anomalies with greater precision compared to traditional methods, reducing the risk of human error and enhancing overall product quality.
- Process Standardization: AI ensures that all operators are trained according to uniform criteria, eliminating subjective variations and ensuring inspections are conducted consistently.
- Operational Efficiency: AI can speed up the training process, reducing the time required to certify operators and increasing productivity.
- Adaptability and Continuous Learning: AI models can be continuously updated and improved, allowing operators to learn new techniques and adapt to changes in inspection requirements.
- Cost Reduction: Using AI can lower costs associated with traditional training, such as educational materials, instructors, and downtime.

- Decision Support: AI can provide real-time feedback to operators during training, helping them better understand their performance and improve quickly.

During the testing phase of the Smart Vision Suite tool (shown in Figure 63), the software was used by a junior inspector who was provided with a series of images containing defects. By analysing and recording the operator's responses, real-time feedback was provided, allowing the operator to get immediately acknowledged about any possible mistakes.

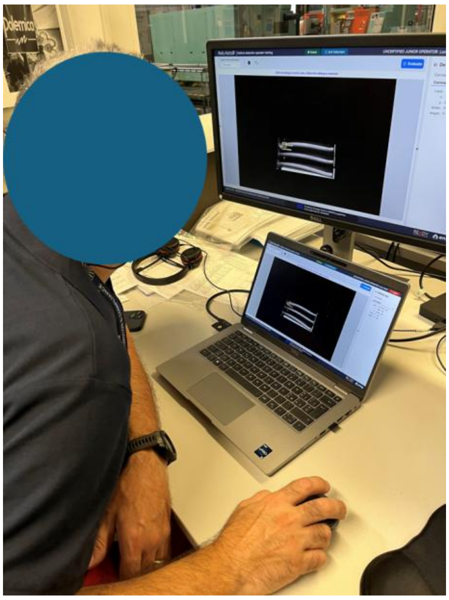


Figure 63 – Operator usage

The primary outcome from the testing phase is that the on-the-job training phase should not be entirely replaced by the Smart Vision Suite tool. This is because a fundamental part of the learning process involves other senses, such as touch, and characteristics that can be better appreciated in person rather than through a picture.

The tool, furthermore, introduces a competency verification methodology that is not currently managed in the existing process and could therefore be effectively applied as:

- Complementary training to on-the-job training: Optimizing training time by providing access to a variety of defects that may not be available on actual parts during the on-the-job training period.
- Competency testing: Introducing a method to certify the skills acquired during the training phase.
- Competency refresh: Establishing a structured method that allows operators to both update their knowledge on new defects that may arise from a quality perspective and verify that their competency level remains consistent over time.

3.a.3.2 KPI Measurement and Performance Evaluation

The three key performance indicators for Pilots 1 and 2 are reported in Table 7.

Table 7 – Pilot 1 and Pilot 2 KPIs

ID	BUSINESS Indicator	DESCRIPTION	Unit*	Initial value
1	Reduce quality control time on the final product	The use of the software will speed up the quality control process	Minutes dedicated to quality check	16min
2	AI Software recognizes the same defects the operator does	To help to operator the software must has a good reliability	Numbers of defect recognized by the software	NA
3	Reduce the number of trainings hours	Using learning software could simulate higher volume production	Hours needed for training	480

The performance evaluation has been conducted depending on the KPI to be assessed.

For KPIs related to Pilot 1, performance evaluation was conducted using two different approaches.

Regarding the first KPI, which focuses on reducing the quality control time for the final product, timed measurements were carried out on the automated inspection process and compared with the manual process.

For the second KPI, which concerns the accuracy of the AI software in recognizing the same defects as the operator, metrics were defined to assess the accuracy achieved by the AI model in defect recognition.

Specifically, the dataset was divided into 80% training, 10% validation, and 10% testing. The percentage of test images was used to measure the accuracy achieved by the model, using the following metrics:

- Precision: the ability of a classification model to identify only the relevant data points.
- Recall: the ability of a model to find all the relevant cases within a dataset.
- F1-score: the average of precision and recall, measuring the model's predictive performance.

Finally, the confusion matrix for all-defect identification has been evaluated

Regarding the third KPI on reducing the number of training hours, an evaluation was conducted by considering the number of different types of defects observed on a specific Part Number over the past year and relating this value to the total number of possible defects.

By narrowing the analysis to only the types of defects observed in the past year and for which the AI algorithm was trained, starting from the number of hours required in the current certification process for a junior operator, the number of different defect types that were observable was assessed. This data was then correlated with the algorithm's ability to present all the defects observed during the acquisition period and samples bench available in production.

3.a.3.3 Final KPI Assessment and Business Impact

The first KPI evaluation of reducing the quality control time on the final product (Table 8) confirm the value already measured in M18.

Table 8 – Pilot 1 and Pilot 2 Final KPIs

ID	BUSINESS Indicator	DESCRIPTION	Unit*	Initial value	Expected value	Expected date of achievement **	Current KPI assessment
1	Reduce quality control time on the final product	The use of the software will speed up the quality control process	Minutes dedicated to quality check	16min	-10%	End of implementation	9min (-44%)
2	AI Software recognizes the same defects the operator does	To help to operator the software must has a good reliability	Numbers of defect recognized by the software	NA	85%	Before 12 months after the implementation	30%
3	Reduce the number of trainings hours	Using learning software could simulate higher volume production	Hours needed for training	480	-10%	End of implementation	330h (-25%)

The previous manual process included four steps:

1. Taking the part (1 minute).
2. Checking the serial number (2 minutes).
3. Conducting a manual visual inspection of the part (10 minutes).
4. Reporting the findings (3 minutes).

This totalled 16 minutes for the manual inspection process of each part.

The implemented automated process follows the same initial steps of taking the part. But then reduces the inspection time by utilizing a RoboCam to capture pictures and applies an AI algorithm to detect defects. The new automated process includes these four steps:

1. Taking the part (1 minute).
2. Checking the serial number (2 minutes).
3. RoboCam and AI algorithm defect recognition (3 minutes).
4. Reporting the findings (3 minutes).

The new process brings the total process time to 9 minutes. This represents a 44% reduction in time from the manual process, which is highly satisfactory compared to our expected 10%-time reduction.

The second KPI, "AI software recognizes the same defects the operator does," has been measured by evaluating the metrics calculated for all-defect detection model in Section 3.a.1.2.

Assuming all defects known in the dataset have been classified as ground truth by the inspectors, as anticipated, only 20% of predicted bounding boxes by model were correct, leading to a high number of false positives. The model detected just 36% of the actual defects, due the complexity of detecting multiple defect types.

The quality of pictures in data collection should improve this result, increasing the reliability of the software to help the operator.

The third KPI on reduction of number of training hours has been measured analysing the number of visual defects detected during last year operations. An assumption has been made, considering the outcomes gave from the senior operators during the test phase: to consider a combination between training provided by the tool and training on-the-job (depicted Figure 64).

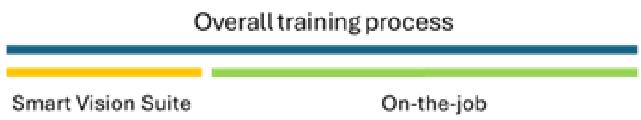


Figure 64 – Overall training process

Based on last year and examining the types of defects that occurred over a random 480-hour timeslot (for example last part of the timeline), it was observed that only 4 out of 7 defect types were encountered during manual inspection phases. By utilizing the AI Smart Vision Suite tool – which has been trained on dataset that included all 7 defect types – (occurred throughout the year), the junior operator would be able to recognize all the defects in the same training period that would otherwise have been encountered over the entire year. This approach optimizes the training process not only in terms of the hours spent to train in recognizing defects but also in terms of completeness, as all defect types identified by the model would be available.

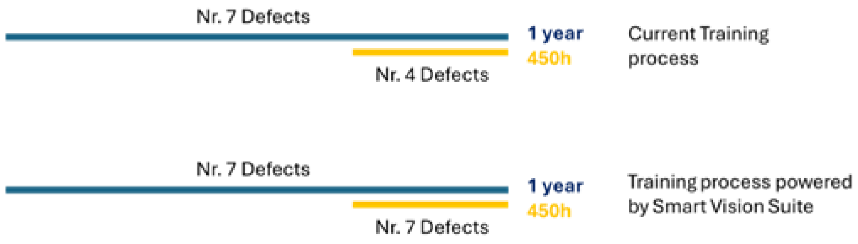


Figure 65 – Training process comparison

From the comparison showed in Figure 65, it is evident that current training process needed to be divided into multiple phases throughout the year in order to observe all the defects identified and documented by quality procedures. With the same number of training hours, however, the Smart Vision Suite tool would allow all defects observed during the reference year to be reviewed.

By dividing the training process into two phases, the phase managed with the Smart Vision Suite software enables a reduction of approximately 120 hours, as showed in Figure 66, presenting all observed defect types and decoupling the training from the availability of parts and, consequently, defects. The estimated reduction in hours was calculated based on the number of inspected part volumes related to production planning and the frequency of defects detected in the manual process.

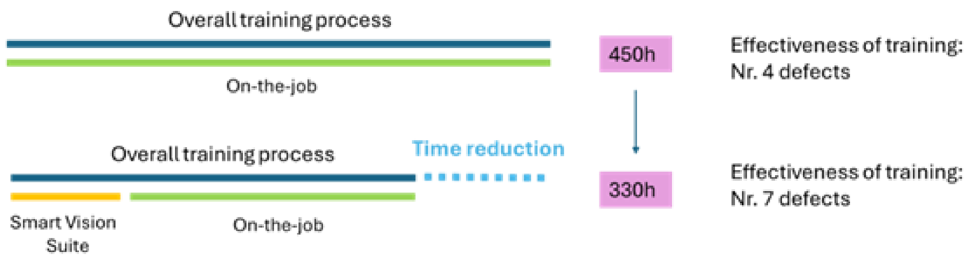


Figure 66 – Time reduction in Training process

Further optimizations could be considered if the tool were used for final and/or periodic tests to verify the competencies acquired by operators.

Additionally, the reporting capabilities allow the training phase to be decoupled from the availability of senior inspectors, thereby optimizing the planning of training sessions.

However, this process remains simulated or, at most, combined with on-the-job training, as current Airworthiness regulations do not yet permit the introduction of AI-based tools for certifying an operator's competencies in the inspection domain.

3.b Business Scenario 3

3.b.1 Full-scale implementation

3.b.1.1 Architecture

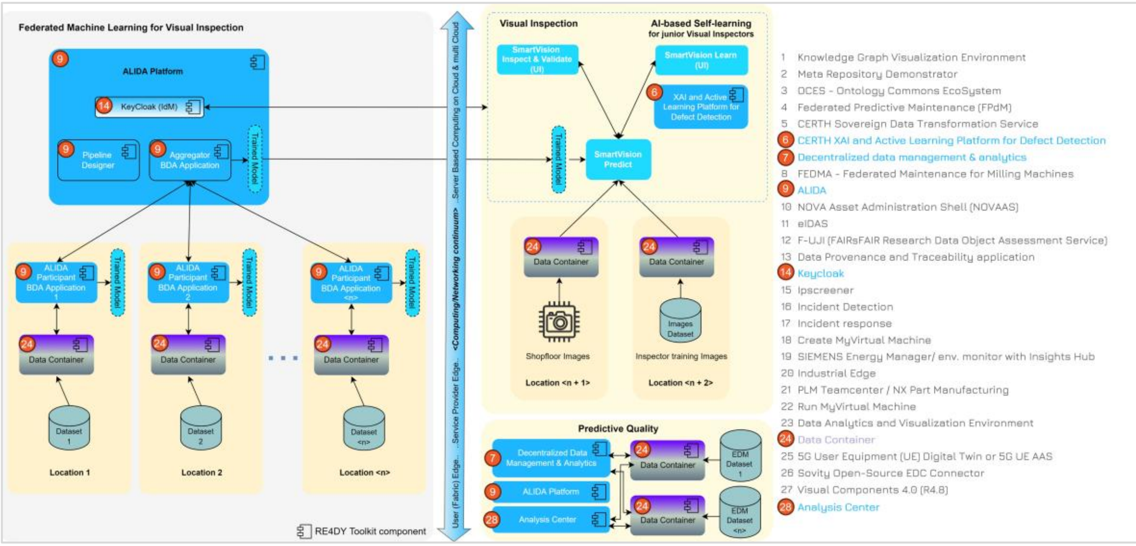


Figure 67 – Implementation of the RA within the AVIO Aero pilot

Similarly to business cases 1 and 2, the toolkit components selected for business case 3 (Figure 67) revealed themselves capable of effectively solving the planned tasks. In particular, business case 3 leverages:

- Component 7: Decentralized data management & analytics
- Component 9: ALIDA
- Component 14: KeyCloak
- Component 24: Data Container
- Component 28: Analysis Center

Components 7 and 28 aim at solving the predictive quality tasks through AI techniques. Component 9 – ALIDA – helps building and deploying pipelines which use AI models from components 7 and 28. Component 14 – KeyCloak – integrates with the project's SSO to provide access to ALIDA. The AVIO AERO-specific implementation of Data Container (Component 24), consists of three modules: S3 Mountpoint, Dataset Aggregator and Scheduler (Figure 68) that work together to provide access to the data source containing the EDM data to be processed by components 7 and 28.

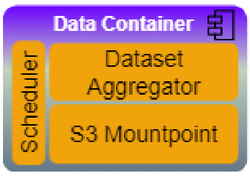


Figure 68 – AVIO Aero-specific Implementation of Data Container

These components (more details in the next sections) fit into the deployment architecture of Figure 69 below. The architecture has remained unchanged, proving functional to achieving its objectives.

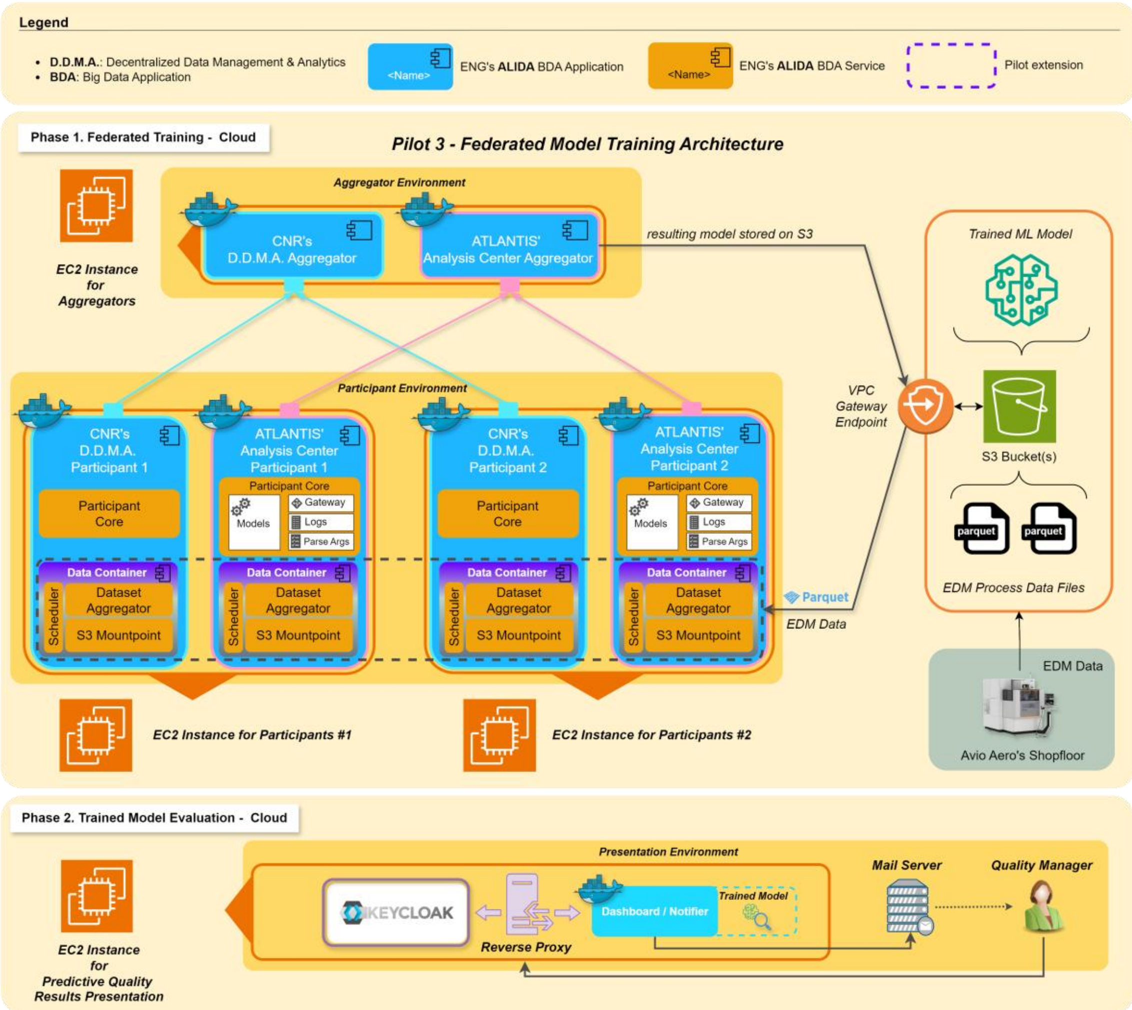


Figure 69 – Deployment architecture for AVIO Aero Business Scenario 3

3.b.1.1.1 Analysis Center (No 28) Architecture

Analysis Center, Number 28 of the Reference Architecture, as seen in Figure 67 is an analytics component developed by Atlantis Engineering SA, able to support monitoring and improvement of quality processes at AVIO AERO. The analytics algorithms were trained based on data obtained from AVIO's EDM Machines and were integrated into the ALIDA Federated Framework, thus offering the end user with a complete and unified solution. A

Grafana dashboard was designed and set up in collaboration with the end user, used for verification of the models and tool validation.

During the trials, the deployment of the core architectural components of the Analysis Center includes the services in Participant, Aggregator, and Presentation environments (see yellow-highlighted areas in deployment architecture – Figure 69)

More precisely, the EC2 machines in the Participant environment emulate the behaviour of industrial assets deployed at distinct production sites – specifically, the Bielsko (PL) and Pomigliano (IT) plants. Each environment hosts localized analytics components that have been registered as ALIDA BDA applications and deployed as containerized applications. In alignment with Federated Learning principles, each Participant is granted access only to the data segments that were generated from its associated site. This data isolation enforces strict privacy and accessibility constraints, and at the same time increases the general knowledge of the final model.

The Aggregator environment orchestrates the Federated Learning process by collecting model updates and metadata from the Participants. The service designed for this purpose is included in the analytics component and has also been registered as an ALIDA BDA application. During deployment, and to follow the Federated Learning principles, the aggregator is deployed in an isolated environment that does not have access to machining data but can only communicate and receive model updates from the Participants.

Finally, the Presentation environment provides the interface for utilization of the final trained model. The primary function of this environment is to render the model outputs, visualize the derived insights, and communicate actionable information to end users. The end user is expected to evaluate the model outputs from the interface hosted in this environment and incorporate the tool into their routine operational workflow.

To ensure seamless integration with the existing infrastructure, each service to be deployed is built into a dedicated Docker image, enabling a fully containerized deployment. The following services are deployed across the different environments:

- Participant Environments: Docker images containing the analytic core services of Analysis Center integrated with ALIDA are deployed. Deployment is managed using Docker and Docker Compose, allowing the customization of environment-specific parameters and providing an easy deployment phase. To adapt to the AVIO AERO Cloud infrastructure, each participant communicates with Smartshop infrastructure that stores all the collected data. Through this channel, data is transferred to the participant as Parquet files. Then, by utilizing ALIDA services, the data is transformed to CSV files that are afterwards given as input for preprocessing and analysis to the developed analytics components.
- Aggregator Environment: A Docker image incorporating the analytics component for model weight aggregation is deployed. Docker Compose is again used, enabling users to configure various parameters related to the training process such as the training epochs, the output directories that store the final model parameters and others.
- Presentation Environment: The presentation environment consists of multiple Docker containers, each supporting a specific component of the end-to-end

operational workflow. One container hosts the PostgreSQL database, another runs a Grafana instance for data visualization and user interaction, and a third handles data ingestion and preprocessing. Specifically, the third container needs machining data from Parquet files, performs preprocessing steps, and stores the processed data in the PostgreSQL database. This data is then passed to a pre-trained model for inference, with the resulting outputs also written to the same database. The coordinated operation of these containers enables a fully functional dashboard that visualizes both raw and machining data and model-derived insights.

A successful and functional deployment of the above system requires careful configuration of each service or container, ensuring that all parameters are properly set to facilitate the following:

- Seamless and continuous communication between the Participants and the Aggregator during training.
- Reliable operation of the services in the Presentation Environment, ensuring that the database and dashboards are continuously updated with fresh machining data and model outputs.

3.b.1.1.2 Prediction Pipeline (Presentation Environment)

Once trained, the AI models (more details in the next sections) power the anomalies detector and are integral part of the following pipeline (Figure 70), which allows for the periodic extraction of EDM data from the source as well as anomalies detection and their visualization. The pipeline is used by both ATL and CNR.

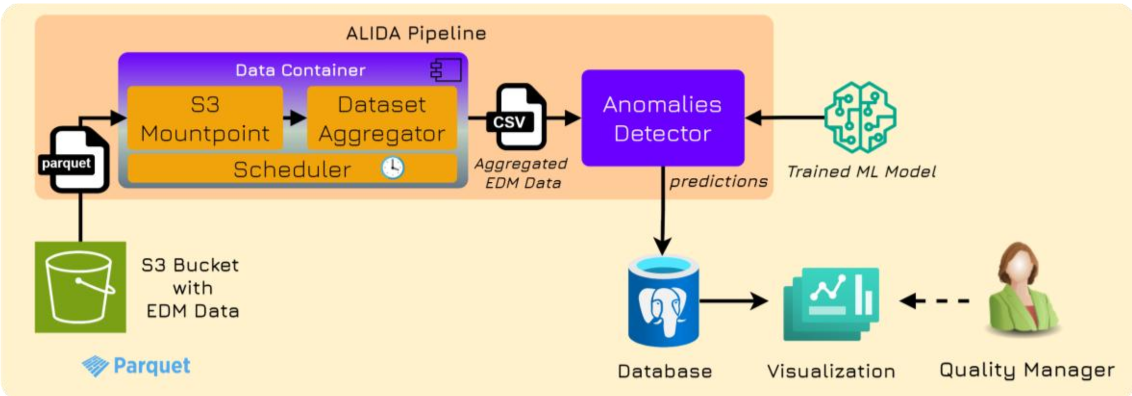


Figure 70 – Prediction workflow used by ATL and CNR

The flow begins with the tri component Data Container. The S3 Mountpoint docker service creates an access channel to the S3 bucket, which contains parquet files with the EDM data. S3 Mountpoint is a tool developed by AWS which, building on top of Linux FUSE (Filesystem in User Space), allows for mounting the buckets onto the local filesystem making their content available via regular directories. At that point, the Dataset Aggregator combines the parquet files into a CSV file, which is finally sent to the Anomalies Detector. The latter, processes the data, detects anomalies and stores them into a database. When the Quality Managers want to visualize the results, they access the visualization tool, which reads the anomalies from the database and presents them in a dashboard. The entire just described pipeline is periodically executed by a scheduler. In addition to that, to allow data scientists with no access to the infrastructure to monitor the status of the running algorithms, a module periodically collects and sends their logs via email.

3.b.1.2 AI Models

3.b.1.2.1 Model Overview

Model architecture: The core of the model is an LSTM (Long Short-Term Memory) Autoencoder (see Figure 71), designed for unsupervised learning tasks. The autoencoder is trained to reconstruct input time-series signals. In this setup, reconstruction error serves as the primary metric for anomaly detection—samples that the model fails to reconstruct accurately (i.e., with high reconstruction error) are flagged as potential anomalies.

Goal: The objective is to identify anomalous behavior in machine-generated time-series data by learning typical signal patterns during a normal operating regime. The model operates in an unsupervised setting, requiring no labeled anomaly data. It relies on the assumption that deviations from learned normal patterns are indicative of abnormal or faulty machine states.

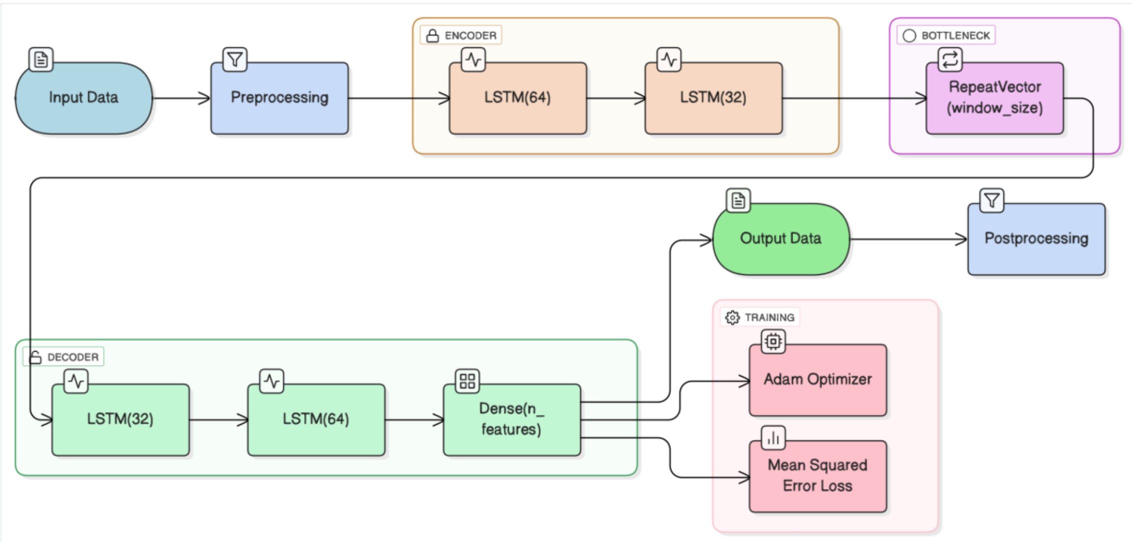


Figure 71 – LSTM-based Autoencoder architecture

Federated Learning Setup: To enhance data privacy and align with decentralized data ownership constraints, a federated learning framework was adopted. Training was distributed across five client nodes:

- *Bielsko Facility*: 2 client instances

- *Pomigliano Facility*: 3 client instances. Each client trains a local model on its own dataset. After each local training iteration, model weights are shared (not the raw data) and aggregated into a global model on a central server. This setup ensures data remains local, maintaining compliance with data protection policies while still benefiting from a collaborative training process.

Training Output Artifacts:

- *Serialized Model File*: Contains the final aggregated LSTM Autoencoder weights, ready for inference deployment.
- *Global Threshold File:* Stores the anomaly score threshold used for flagging anomalies. This is computed based on the global reconstruction error distribution, assuming a contamination rate of 5%, i.e., this is a standard assumption in the context of anomaly detection.

3.b.1.2.2 Data Preprocessing

Signal Selection Criteria: From the available sensor data, seven signals were selected for training. Selection was based on three key criteria:

1. High Variance: Ensures that the selected signals carry meaningful dynamics.
2. Low Missing Value Rate: Minimizes the impact of imputation bias and increases reliability.
3. Low Inter-Signal Correlation: Encourages diversity, reducing redundancy across input features.

Missing Data Handling: Various imputation methods were evaluated for handling NaN values. Forward fill (propagating the last observed value) was ultimately chosen due to its neutral impact on signal continuity and overall stability across experiments.
Normalization: All input signals were normalized using standard normalization (z-score) to ensure uniform scaling and accelerate training convergence.

Dataset Scope:

- Machines Included: “A04858”, “A04859”, “A04668”, “A04672”, “A04673”
- Temporal Scope: Only the first available month of data per machine was used for model training to emulate early-stage anomaly learning.
- Windowing Strategy: Time-series data were segmented into fixed-size windows of 300 samples. Multiple window sizes were tested, but 300 samples offered the best balance between context length and granularity.

3.b.1.2.3 Experiment Setup

Federated Training Process: Training was conducted using a federated learning framework, where each client independently trains a local LSTM Autoencoder on its windowed dataset. After local training epochs, model weights are communicated to a central server for aggregation. The model is trained to minimize the reconstruction loss, specifically *Mean Squared Error* (MSE), with no requirement for labeled anomalies.

Anomaly Detection Procedure

Anomaly Score: Calculated as the reconstruction error of the autoencoder for each time-step.

Thresholding: A post-training threshold was established by analyzing the global error distribution and setting the cutoff at the 95th percentile, assuming a contamination rate of 5%, in line with standard practices in unsupervised anomaly detection.

Prediction Granularity: Anomalies are flagged at the individual timestamp level rather than over intervals, ensuring high temporal precision.

3.b.1.2.4 Results & Evaluation

Unsupervised Consistency Check:_The trained model was applied to data from later months to assess stability. Anomaly rates in these validation periods remained close to the expected 5%, confirming that the model generalizes well and does not overfit to the training window.

Supervised Validation (Synthetic Anomalies):_To assess model performance under known anomalous conditions, synthetic perturbations were introduced:

- Noise with ±3σ magnitude was added to approximately ~X% of data points in a subset of the selected signals.
- The model's ability to detect these perturbations was evaluated using supervised metrics such as accuracy and F1-score.
- Results indicated strong detection capability, demonstrating the model's robustness and reliability in practical anomaly detection scenarios.

Figure 72 shows an example of the core outputs of the LSTM Autoencoder applied to a one-hour time interval of a selected signal.

Signal Details: The example focuses on the ARCKILL signal, a key parameter monitored during machine operation. The raw signal is plotted over a continuous one-hour window, providing a clear view of its dynamics.

Overlayed Reconstructions: Alongside the original ARCKILL signal, the corresponding reconstructed signal, named ARCKILL_recon, is superimposed. The LSTM Autoencoder generates this reconstructed output and represents the model's best attempt to replicate the original time-series based on patterns it has learned during training. The visual comparison between the original and reconstructed signal helps highlight any significant deviations, which are potential indicators of anomalous behavior.

Anomaly Score Plot: Beneath the signal plots, the anomaly score is charted across the same time axis. The anomaly score is computed as the reconstruction error—typically the Mean Squared Error (MSE) between the input and output vectors for each step. Elevated anomaly scores suggest a poor reconstruction by the model, and therefore a higher likelihood of anomalous behavior at that timestamp.

Binary Anomaly Label: To enhance interpretability, a binary anomaly label is also included in the plot. This label takes a value of 1 when the anomaly score exceeds a predefined global threshold (calculated during the model evaluation phase based on an assumed contamination rate), and zero otherwise. The label provides a straightforward, threshold-based classification of time points as anomalous or normal, supporting downstream analysis or alerting mechanisms.

The second visualization is provided to complement the anomaly detection output. This graph spans the same one-hour interval and reports a *smoothed average* of the signal or anomaly score, computed over a sliding window or rolling mean.

The purpose of this graph is to:

- Offer a noise-reduced view of signal behavior, which can help identify gradual trends or subtle shifts not immediately obvious in raw signal data.
- Enhance the visual clarity of underlying patterns and improve the analyst's ability to correlate anomalies with contextual changes in the signal.

Figure *72* – Anomaly detection output example in two separate visualisations. On the left, the signal is deemed as anomalous and, on the right, the specific anomalous pattern is accompanied by a numerical indication of the anomaly score

3.b.1.2.5 AI Models results of Analysis Center

Model

The core component of the Analysis Center is a fully connected Autoencoder, designed specifically for this use case. Given the unavailability of labelled data, an unsupervised modelling approach was adopted. The Autoencoder consists of two primary components: an encoder and a decoder. The encoder compresses the input time series data into a lower-dimensional latent representation, capturing the essential patterns of normal operational behaviour of the EDM machines. The decoder subsequently reconstructs the input data from this latent representation. This architecture allows the model to learn a compact and representative encoding of healthy operational states, serving as the basis for anomaly detection.

Preprocessing

The data preprocessing pipeline was specifically designed to accommodate the characteristics of the EDM datasets provided. The datasets originated from two sites and encompassed a total of five machines, each exhibiting distinct operational behaviours.

These differences necessitated tailored preprocessing procedures to ensure consistent representation across machines.

Initial preprocessing included standard data cleaning techniques, such as the removal of noise and handling of missing values. Afterwards, the data were synchronized in order to obtain a unified dataset with the same intervals between all available timeseries. However, the most critical step involved organizing the data into meaningful time windows that preserved the intrinsic operational structure. After detailed examination, the data were segmented into smaller units referred to as sequences. Each sequence represents a small-time window during which the machine operates on a specific piece of equipment.

For each sequence, statistical features—including mean, standard deviation, and lagged values—were extracted from non-static features to capture the temporal and operational characteristics of the machine. These sequences, along with their derived features, were then used as input to the Autoencoder for anomaly detection, where each sequence could be classified as either healthy or anomalous.

The utilized features were primarily related to electric spark behaviour, as well as operational and contextual metrics. The complete set of features used as input to the model is in Table 9:

Table 9 – Complete set of features used as input to the model

Feature	Description
PARTNUM	Part number
TOOLNUM	Tool number
EFFICIENCY	Average efficiency of the current setting (%)
ARCVOLTAGE	Bad sparks due to arc voltage by IPG generator (%)
DELAY	Bad sparks due to delay by IPG generator (%)
SEQUENCETIME	Sequence time
GOOD	Good sparks by IPG generator (%)
MACHININGTIME	Machining time
ARCKILL	Actions taken on bad sparks killed by IPG generator (%)
SHORTCIRCUIT	Percentage of short-circuit sparks
MACHININGSPEED	Machining speed
SEQUENCE	Sequence number
ESTOP	Emergency stop
PARTJOBNAME	Job name
EXECUTION	Execution coding (e.g. 'Active', 'Stopped', 'Interrupted', etc.)

Training & Testing

During training, the Autoencoder is exposed exclusively to time series slices corresponding to normal machine operation. The model learns to recreate these healthy slices with minimal reconstruction error, effectively capturing the intrinsic dynamics and correlations present in the data. Optimization is performed to minimize the reconstruction loss, which serves as a measure of the difference between the original input and its reconstructed output. By focusing solely on healthy data, the model inherently becomes sensitive to deviations from normal behaviour, without requiring labelled examples of faults or anomalies. The split used during training was 70/15/15 for train/validation/test set accordingly.

For evaluation, new time series slices are fed into the trained Autoencoder. When the input corresponds to healthy operation, the model is expected to reproduce the data with low reconstruction error. Conversely, when a slice exhibits abnormal or faulty behavior, the reconstruction error increases significantly due to the model's lack of prior exposure to such patterns. To determine whether a slice is anomalous, a threshold was defined based on the reconstruction error as follows:

$$abs\left(\frac{ReconstructionError - Mean(OverallReconstructionError)}{S\tan dardDeviation(OverallReconstructionError)}\right) > 3$$

Figure 73 – Reconstruction error formula

This formulation introduces a tolerance factor into the anomaly detection process, allowing the method to account for natural variations in normal operational data. Slices exceeding this threshold are classified as anomalous, whereas those below the threshold are considered healthy.

Evaluation & Observations

Due to the absence of labelled data, the evaluation of the Autoencoder was conducted using manually generated anomalies. These synthetic anomalies were created by systematically increasing the values of features indicative of faulty operation, such as ARCKILL, ARCVOLTAGE, and SHORTCIRCUIT, while simultaneously decreasing the values of features representing normal operation, including GOOD and EFFICIENCY. This approach allowed for controlled testing of the model's sensitivity to abnormal behaviour. Additionally, we observed that model performance improved when training was restricted to machines with similar operational characteristics. In particular, one machine from a specific site (Pomigliano site, machine A04668) exhibited operational patterns that were substantially different from the other machines, and including its data during training reduced the model's ability to generalize. In future work, the model should also be evaluated against anomalies that have been manually labelled or recognized by operators to ensure alignment with real-world fault detection.

3.b.1.3 Applications

The Analysis Center can be conceptualized as an application composed of two primary components: the model-related module and the user interface.

The model-related module is responsible for processing EDM data and producing anomaly predictions. It accepts input in the form of CSV files, which are passed through a custom preprocessing pipeline before being fed into the trained Autoencoder. The module outputs a classification for each sequence, labelling it as either healthy or anomalous. All inference results are stored in a database, allowing subsequent access by users or visualization tools. Additionally, the module incorporates a comprehensive logging system, which records all processing steps, facilitating transparency and traceability.

The user interface provides a means for users to monitor the AI system and interpret its results through interactive tools, including:

- Interactive Dashboard: Grafana dashboards were developed to visually represent model outputs. As illustrated in Figure 74 and Figure 75, the dashboard includes both a table and a graphical panel. The table provides the following information for each evaluated sequence:
- Machine: A unique identifier for the machine corresponding to the inference.
- Inference: The anomaly score assigned to the sequence. The column is color-coded to indicate severity: low scores in green, medium scores in yellow, and high scores in red.
- Sequence Start: Timestamp marking the beginning of the data sequence.
- Sequence End: Timestamp marking the end of the data sequence.
- Sequence: A hyperlink enabling the user to isolate and examine the data for a specific sequence, facilitating focused analysis of individual cases (Figure 75).

Figure 74 – Interactive dashboard for analytics visualization

Figure 75 – Interactive dashboard for a specific sequence

Logging System

Each component of the Analysis Center logs its operations to provide transparency and ensure robust monitoring. Logs include details on input files being processed, the anomaly scores computed for each sequence, and confirmation of completed processing, indicating readiness for subsequent files. The logging system also captures errors, such as unsupported file formats, thereby supporting efficient troubleshooting and operational reliability.

```
predictor-1  | [2025-05-12 13:44:10.417] Checking for existing CSV files in /data ...
predictor-1  | [2025-05-12 13:44:16.603] INFO:inference Processing file /data/A04858.csv
predictor-1  | [2025-05-12 13:44:17.284] INFO:inference Processing data for machine M_A04858
predictor-1  | [2025-05-12 13:45:18.347] INFO:inference Data successfully written to 'postgres:machines'.
predictor-1  | [2025-05-12 13:45:42.504] INFO:inference Data successfully written to 'postgres:edm_data'.
predictor-1  | [2025-05-12 13:46:48.929] INFO:flower Average Root Mean Squared Error between original/reconstructed arrays: 6.030760451508466
predictor-1  | [2025-05-12 13:46:49.202] INFO:inference Data successfully written to 'postgres:inferences'.
predictor-1  | [2025-05-12 13:46:49.208] INFO:inference Machine M_A04858 - Sequence From 2024-09-02 05:28:08 To 2024-09-02 05:46:41 - Anomaly: False
predictor-1  | [2025-05-12 13:46:49.208] INFO:inference Machine M_A04858 - Sequence From 2024-09-02 05:46:41 To 2024-09-02 05:53:26 - Anomaly: False
predictor-1  | [2025-05-12 13:46:49.209] INFO:inference Machine M_A04858 - Sequence From 2024-09-02 06:58:31 To 2024-09-02 07:00:02 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.925] INFO:inference Data successfully written to 'postgres:inferences'.
predictor-1  | [2025-05-12 13:47:08.927] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 22:42:16 To 2024-09-25 22:48:08 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.928] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 22:48:14 To 2024-09-25 22:54:52 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.929] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 22:54:57 To 2024-09-25 22:56:05 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.929] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 22:56:10 To 2024-09-25 23:03:42 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.930] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 23:03:44 To 2024-09-25 23:21:12 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.931] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 23:21:18 To 2024-09-25 23:27:58 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.932] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 23:28:00 To 2024-09-25 23:28:57 - Anomaly: False
predictor-1  | [2025-05-12 13:47:08.933] INFO:inference Machine M_A04859 - Sequence From 2024-09-25 23:29:02 To 2024-09-25 23:31:14 - Anomaly: False
predictor-1  | [2025-05-12 13:47:09.389] Watching for new CSV files in /data ...
```

Figure 76 – Presentation environment logging mechanism

3.b.1.4 Key challenges and solutions for full-scale implementation

3.b.1.4.1 Challenges of Analysis Center

Although the developed component is fully functional, an additional effort has been made to address a couple of challenges in order to improve the scalability of the implementation:

- Participants: In a full-scale deployment, participants should be hosted on the edge – directly on or near the machinery equipment – rather than in cloud-based environments. This edge deployment would eliminate unnecessary data transfer over external networks by enabling local processing, thus reducing latency and potential security risks. Additionally, it is essential to ensure that each edge device has sufficient computational and memory resources to handle the potentially intensive workload required by federated learning tasks, which may vary depending on the final goal and model complexity.
- Aggregator: As the number of participants increases, the aggregator must communicate efficiently with all edge nodes to ensure timely model updates. While secure channels are already in place from the utilized federated learning framework (Flower framework uses gRPC channels), maintaining high-speed communication becomes a key scalability challenge. To address this, the system should minimize synchronization delays by optimizing data transfer protocols, reducing payload sizes, and possibly employing parallel or asynchronous aggregation strategies.

In addition to participant and aggregator-oriented improvements, a full-scale implementation would require continuous monitoring of edge participants. This includes detecting performance bottlenecks, identifying node failures, and addressing connectivity issues. Robust observability mechanisms should be implemented, including real-time health checks, resource usage tracking, and automated alerting.

Beyond monitoring, the system should also include fault-tolerant mechanisms to handle edge node failures smoothly and without disrupting the overall federated learning process. This could involve strategies such as temporarily excluding unresponsive nodes, rescheduling training rounds, or using asynchronous updates to maintain network stability and model convergence.

3.b.2 Industrial trials of the pilot

3.b.2.1 Testing procedure and Barriers

The main barrier encountered while deploying and testing the tools on the Avio Aero IT infrastructure was basically the same of previous pilots: the need to be compliant with the strict cyber security regulations. Access to the virtual machines and data storage could only occur from within the Avio Aero network and with Avio Aero-compliant computers. From the laptop with personal SSO credentials was possible to access the Avio Aero network through a VPN and to the specific virtual machines through CyberArk.

3.b.3 Final KPIs monitoring and validation

3.b.3.1 Industrial Outcomes and Lessons Learned

The Industrial Pilot implementation for Business Scenario 3, assumes different business key factors:

- Cost Reduction: Predictive quality minimizes costs associated with defects, rework, and warranty claims by identifying issues early in the process.
- Efficiency Gains: Optimizing production processes and resource allocation through predictive insights leads to improved operational efficiency.
- Risk Mitigation: Early detection of quality issues reduces risks related to product recalls, compliance violations, and customer dissatisfaction.
- Competitive Advantage: Implementing predictive quality enhances product reliability and consistency, differentiating the business in the market.
- Customer Satisfaction: Delivering high-quality products consistently builds customer trust and loyalty, driving repeat business and positive brand reputation.
- Data Utilization: Leveraging advanced analytics and machine learning to extract actionable insights from production and quality data maximizes the value of existing data assets.
- Scalability: Predictive quality systems can be scaled across multiple production lines or facilities, ensuring consistent quality standards globally.

Based on models results, an outcome was the low variety of signals in dataset exposed by the equipment which is translated in a not exhaustive description of the overall process.

The upgrade of software version embedded on the equipment would enrich the signals, but this implies the certification of the entire process with extra costs and long times for the business.

In addition, despite that EDM process between equipment is provided on the same product, it could be related to different operations in production cycle based on Part Number: this could be a limitation in terms of predicted parameters process configuration around all equipment. This limitation is strictly related to the production management in Avio Aero, having a logic model of production line: an equipment works different Part Numbers, and the production cycle can consider the same equipment for different cycle operations depending by the Part Number. The lesson learnt is that to maximize the effectiveness of federated prediction models, equipment that provide same operations across the sites must be considered.

An additional outcome is the unavailability of direct timestamp correlation between process deviations coming from equipment signals and product quality during inspection processes, that happen later in production cycle. This is translated in missing connection between predicted quality feedback coming from models and quality losses contribution in Overall Effectiveness Equipment calculation. This not depending by the pilot implementation, but it is strictly connected to the overall process that does not include the quality inspection at the same time of the EDM operation. If in the next future a traceability correlation will be introduced across different operations, so between signals from equipment and quality inspection in this case, the quality contribution can be counted in the OEE and productivity of the equipment.

3.b.3.2 KPI Measurement and Performance Evaluation

The key performance indicator for Pilot 3 is reported in Table 10.

Table 10 – Pilot 3 KPIs

ID	BUSINESS Indicators	DESCRIPTION	Unit*	Initial value	M18 Value	Expected value	Expected date of achievement**
1	Efficiency of OEE on EDM machines.	Data collection could implement both Predictive Quality and Predictive Maintenance by intervening on 2 of the 3 indices that calculates OEE	OEE	75%	TBA	+1%	Before 24 months after the implementation

The Performance evaluation has been addressed through the OEE calculation based on data collected on last year. Additionally, to the signals managed by the AI models developed and implemented for this Pilot, the signal on the machine status (exec) has been collected. The value assumed by the signal is than classified as uptime or downtime in OEE calculation. More details are listed in Table 11:

Table 11 – Signal details

Signal	Description	Unity
exec	State of Machine	READY, ACTIVE, INTERRUPTED, STOPPED

OEE is expressed as a percentage and is calculated by combining three factors:

- Availability: The percentage of scheduled time that the equipment is available to operate.
- Performance: The speed at which the equipment operates compared to its designed speed.
- Quality: The percentage of good parts produced compared to the total parts produced.

The OEE calculation splits the contribution loss depending by the classification of different values of signals. The OEE formula is:

OEE = Availability x Performance x Quality

Where:

Availability = Operating Time / Scheduled Time
Performance = (Ideal Cycle Time x Total Parts) / Operating Time
Quality = Good Parts / Total Parts

Table 12 lists the classification adopted for Machine Status signal:

Table 12 – Machine Status Signal

Uptime	Downtime (Availability/Performance)
ACTIVE	READY, INTERRUPTED, STOPPED

An additional signal related to Cycle Time is used for Performances losses calculation.

As shown in Table 12, the downtimes for EDM equipment include only Availability and Performances losses, assuming there are no quality losses. This is a strong assumption due the restricted availability of data and for the missing direct correlation between status signal and quality inspection (as after operation in the production cycle).

3.b.3.3 Final KPI Assessment and Business Impact

Compared to the OEE values calculated at the beginning of the project, focusing the performance analysis on the last year, the initial OEE values for each equipment (both for the Pomigliano and Bielsko facilities) are significantly lower. This was due to a decrease in production volumes, resulting in reduced utilization in terms of the capacity of the selected machines. Therefore, the OEE percentage increase is assessed based on this assumption.

Table 13 depicts the performance analysed for the KPI of pilot 3.

Table 13 – Pilot 3 Final KPIs

ID	BUSINESS Indicators	DESCRIPTION	Unit*	Initial value	Expected value	Expected date of achievement**	Current KPI Assessment
1	Efficiency of OEE on EDM machines.	Data collection could implement both Predictive Quality and Predictive Maintenance by intervening on 2 of the 3 indices that calculates OEE	OEE	75% (14%-45% in 2025)	+1%	Before 24 months after the implementation	11%-73% (+1÷20%)

The OEE measured at the beginning was higher than the value in 2025, depending by the contribution in Availability and Performances losses. The current value is lower: 14%-45% in 2025, depending by the equipment). The expected OEE value has been evaluated in terms of absolute increase, as already defined in Table 13.

The process engineer analysed the feedback provided by the algorithms using the dashboards developed by Atlantis and CNR showed in Figure 77, validating the AI models while simultaneously acting on the process when the deviation was confirmed.

Figure 77 – UI Dashboard: Process Monitoring

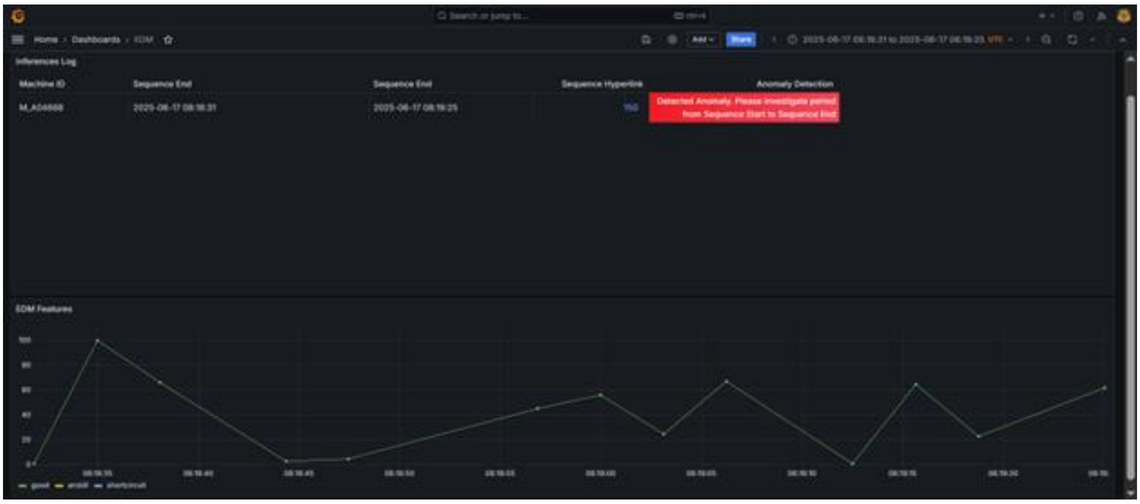


Figure 78 – UI Dashboard: Anomaly Detection

When an anomaly is notified and confirmed by the operator (see Figure 78), the process engineer promptly modifies the process parameters whenever possible, avoiding downtime or (in a still very preliminary phase) a potential impact on the quality of the processing and, consequently, the final product.

By analysing the OEE trend and focusing on the last month (when the tool has been tested by process engineers), it can be observed that the trend is increasing, with an improvement that significantly exceeds 1%, for some equipment.

Below (Figure 79 to Figure 82) are shown the OEE trends and the individual contributions of losses for two different equipment, in Pomigliano and Bielsko respectively.

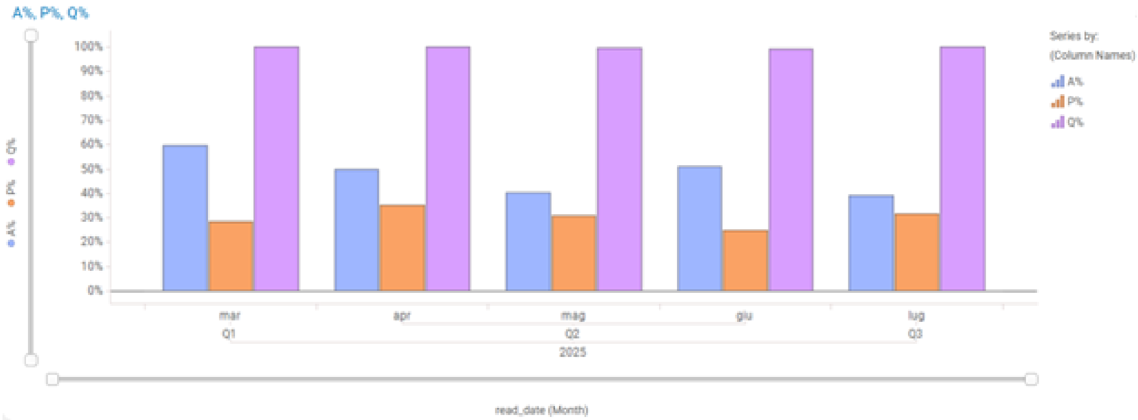


Figure 79 – Equipment A04858 (Pomigliano) - Losses trends

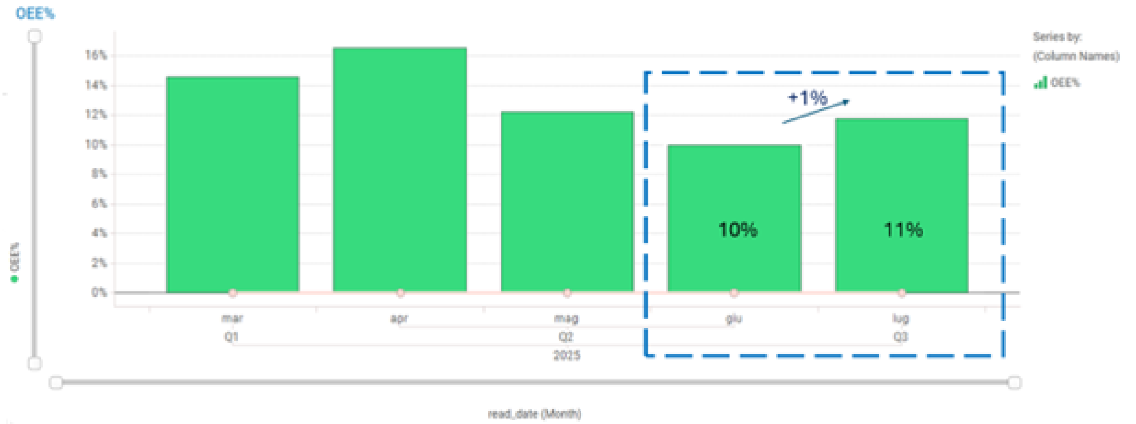


Figure 80 – Equipment A04858 (Pomigliano) – OEE trend

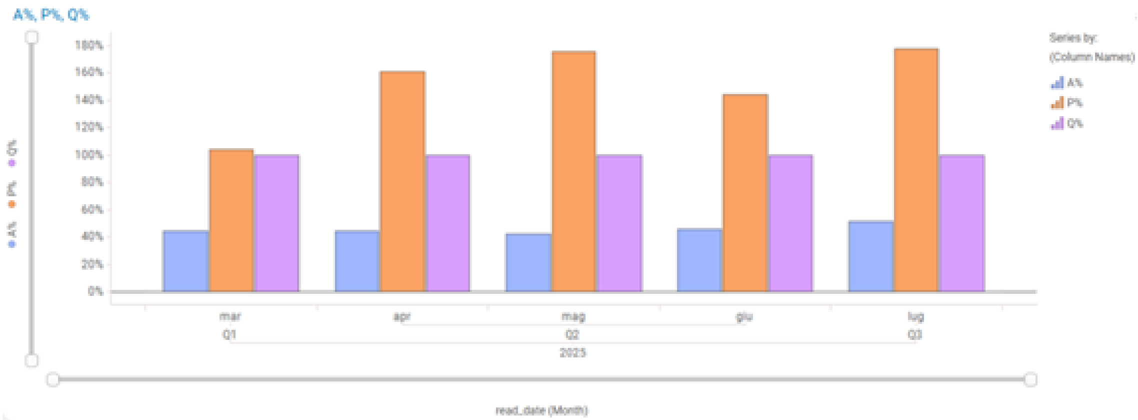


Figure 81 – Equipment A04673 (Bielsko-Biala) - Losses trends

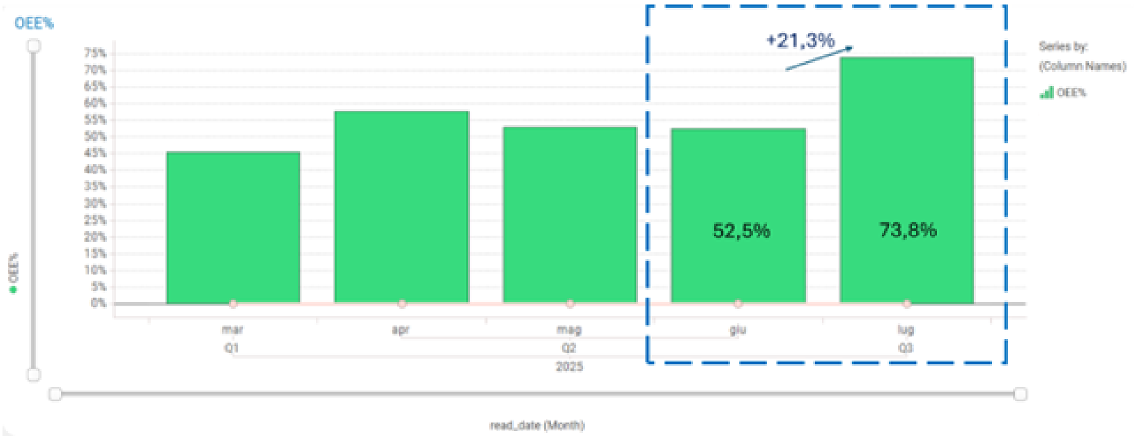


Figure 82 – Equipment A04673 (Bielsko-Biala) – OEE trend

As previously mentioned, the increase in OEE is not directly imputable to an impact on quality, as it has been assumed that quality is always ensured without contributions from quality losses. Process optimizations and production increases have impacted on the contributions of Availability and Performance. The management of AI model feedback on the process may have contributed to reducing and optimizing losses, thereby impacting the increase in OEE value.

The company's current objective is to work on the traceability of the parts processed within the production cycle in order to properly manage the contribution of quality losses in the OEE calculation, achieving a value that is more realistic and closer to the actual process. In this way, the AI models used, in addition to impacting Availability and Performance losses for the reasons described above, will have a direct impact on optimizing the contribution of Quality losses, improving both the efficiency of the plant and the quality of the process itself.

4 Performance Monitoring Framework

4.1 Introduction of methodology (6P-Performance Pillar)

The 6Ps Migration Model functions as a strategic framework designed to support organization particularly those in the manufacturing sector in assessing both their current level of digitalization and their target maturity across six key dimensions. Additionally, it offers a structured means to monitor progress along a digital transformation roadmap[5].

When an organization becomes aware of its digitalization gaps, two main scenarios may arise:

- In the first, the organization is already involved in a project that targets specific aspects of its digital strategy. In such cases, the implementation roadmap is often shaped by the project's scope, and the 6Ps model provides a mechanism to evaluate the progress and impact of the ongoing initiatives.
- In the second scenario, the organization has recognized the gaps but has yet to address them. Here, the model helps by identifying areas for potential improvement, thus guiding strategic planning.

To fulfill its purpose, the model performs a comprehensive analysis of six core pillars that reflect essential components of the production environment. It is built on the understanding that successful digital transformation must go beyond technical capabilities and include what are referred to as "socio-business" dimensions. The six pillars—Product, Process, Platform, People, Partnership, and Performance—are grouped into three technical and three socio-business categories, offering a balanced and integrated perspective for evaluating digital transformation efforts, as Figure 83.

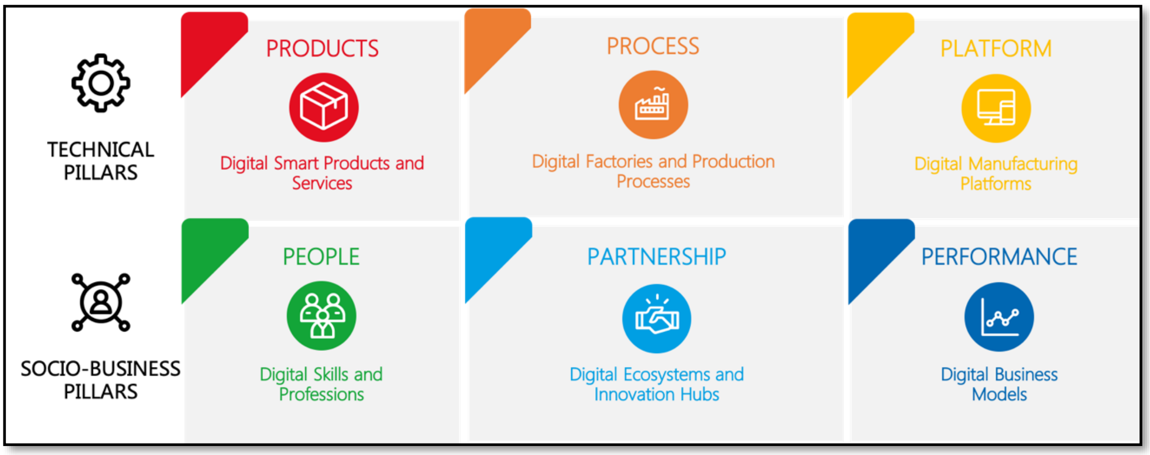


Figure 83 – The 6Ps Model

5 https://re.public.polimi.it/retrieve/handle/11311/1206091/715004/1-s2.0-S2212827122001068-main.pdf

Within the scope of this work package, particularly Task 5.4, titled "Pilot Area KPI Collection and Benchmarking Data", the focus is placed on the Performance dimension of the 6Ps model. This pillar plays a critical role in examining how indicators within manufacturing environments are defined, measured, and monitored. Importantly, the emphasis of this dimension is not on whether indicator values themselves have improved, but rather on the extent to which measurement practices have become more precise and reliable. The Performance dimension is structured into six interrelated areas: Operational/Technical, Economic, Environmental, Social, Product-Service Lifecycle, and Supply Chain. Together, these areas offer a comprehensive lens through which performance-related digital capabilities can be assessed. To determine a company's level of digital maturity, responses to the assessment are mapped onto a structured five-level maturity scale, enabling a standardized and comparative evaluation across different pilot sites.

- Initial: in this stage, the dimension is poorly digitized or not digitalized at all. Processes are poorly controlled, if at all, and managed reactively.
- Managed: at this level, some aspects of the organization are digitalized and controlled, such as through a pilot or an ongoing digitization project. Processes are partially controlled and managed based on experience.
- Defined: in this stage, digitalized activities are defined and implemented throughout the organization. Processes are planned and adhere to good practices and management procedures.
- Integrated: processes are fully planned and implemented, with a focus on information exchange, integration, and interoperability across applications. Best practices and common standards are present.
- Exploited: at this highest level, the organization fully exploits the dimensions. Processes are digitally oriented and built upon a robust technology infrastructure. The organization demonstrates high potential for growth and supports decision-making effectively.

Industry 4.0

	LEVEL-1 INITIAL	LEVEL-2 MANAGED	LEVEL-3 DEFINED	LEVEL-4 INTEGRATED	LEVEL-5 EXPLOITED
OPERATIONAL/ TECHNICAL	Operational performance is often not measured or understood	Descriptive Performance - Measurement and analysis of business KPIs are largely retrospective	Diagnostic Performance - Measurement of KPIs is clear. Attempt to understand the causes that affects events and behaviours	Predictive Performance - Measurement of KPIs is prospective. Statistical models and forecasts techniques to understand the future KPIs	Prescriptive Performance – future-oriented. Optimization and simulation to find the best course of action and operational KPIs measurement
ECONOMIC	Economic performance is often not measured or understood	Descriptive – Measurement of economic KPIs is largely retrospective	Diagnostic - Measurement of economic KPIs is clear. Attempt to understand the causes of events and behaviours	Predictive - Measurement of economic KPIs is prospective. Statistical models and forecasts techniques to understand the future	Prescriptive – future-oriented. Optimization and simulation to find the best course of action and economic KPIs measurement
ENVIRONMENTAL	Environmental performance is often not measured or understood	Descriptive – Measurement of environmental KPIs is largely retrospective	Diagnostic - Measurement of environmental KPIs is clear. Attempt to understand the causes of events and behaviours	Predictive - Measurement of environmental KPIs is prospective. Statistical models and forecasts techniques to understand the future	Prescriptive – future-oriented. Optimization and simulation to find the best course of action and environmental KPIs measurement
SOCIAL	Social performance is often not measured or understood	Descriptive - Measurement of social KPIs is largely retrospective	Diagnostic - Measurement of social KPIs is clear. Attempt to understand the causes of events and behaviours	Predictive - Measurement of social KPIs is prospective. Statistical models and forecasts techniques to understand the future	Prescriptive – future-oriented. Optimization and simulation to find the best course of action and social KPIs measurement
PRODUCT-SERVICE LIFECYCLE	No product life cycle assessment	A few life-cycle aspects are included in some KPIs but occasionally	Life Cycle Costing (LCC) towards recycling, de-re-manufacturing KPIs	Life Cycle Costing + Environmental LCA towards Circular Economy	Life Cycle Costing + Environmental LCA + Social LCA towards Sustainability and Green Deal
SUPPLY CHAIN	Performance is often not measured or understood	Only the most important physical performance of suppliers (e.g. punctuality, quality, operational flexibility)	Physical and Economical performance (purchase price, non-quality costs, delivery delays, lack of flexibility, etc.).	Physical, economical, sustainability performance for almost all the suppliers.	Physical, economical, sustainability and integration with other external sources (e.g., social media, weather)

Figure 84 – 6Ps Model – Performance dimension

Given that the 6Ps method serves as a comprehensive monitoring framework, its structured maturity levels enable a standardized and replicable assessment across multiple pilot sites. This is particularly valuable for supporting the generalization of impact KPIs beyond isolated factory-level insights. The Performance dimension, in particular, provides a robust basis for evaluation, as it encompasses a wide range of critical areas (shown in Figure 84). This comprehensive scope aligns well with the objectives of the task 5.4, which aims to facilitate a generalized and integrated impact assessment across the entire life cycle of the value network.

Moreover, the flexibility of the 6Ps model makes it especially suitable for incorporating technical elements within the analysis. By integrating components such as Process Planning and Preparation, Data Sharing and Integration, Federated Learning and AI Models, and Progress in Integration, the model is also aligned with the technical goals of the work package. This integration not only enhances visibility and measurement but also strengthens the framework's capacity to monitor production performance and effectively generalize impact KPIs across diverse contexts.

In this regard, the newly analyzed aspects have been formulated and structured as it follows:

- Pilot business processes: The primary objective of this section was to evaluate the current status and effectiveness of process planning and preparation activities within the pilot implementations. By asking respondents to rate progress, identify achieved milestones, and outline encountered challenges.

- Data Sharing and Integration: The questions aimed to capture the level of technical progress in connecting and synchronizing data sources, identify specific barriers and limitations encountered during implementation, and evaluate the perceived effectiveness of the Data Container in supporting data exchange and digital service enablement. Additionally, the survey sought to collect evidence of tangible benefits derived from its use, such as improved transparency, Cost savings from streamlined data processes, and data-driven decision-making capabilities.
- AI Models and Federated Learning: This section aimed to measure the perceived advancement in adoption, assess how effectively Federated Learning has contributed to enhancing AI model performance, and identify tangible benefits such as improved model accuracy, enhanced data privacy, and reduced data transfer costs. It also sought to uncover key barriers to adoption, including technical, organizational, and integration-related challenges, and gather additional qualitative insights through open-ended feedback.
- Progress in Integration: This section aimed to capture both the quantitative status of integration efforts and the specific achievements realized, such as data synchronization, digital twin development, and predictive maintenance implementation. It also sought to identify common barriers and operational challenges that may hinder progress, including technical, organizational, or resource-related issues. Finally, the open-ended question was included to collect any additional qualitative feedback from pilot teams.

The methodology is structured around five key steps; each aligned with the principles of the 6Ps framework and tailored to the specific activities of each pilot and the objectives of Task 5.4:

1. Design and preparation of the survey, guided by the 6Ps methodology and adapted to reflect the context and operational realities of each pilot site.
2. Identification of the AS-IS profile, representing the initial digital and organizational maturity of the manufacturing enterprise through structured survey. (beginning of the project)
3. Identification of the TO-BE profile, outlining the targeted or expected future state of the enterprise in terms of digital transformation.
4. Identification of actions to bridge the identified gaps, which may include collaboration with project partners and the continuous monitoring of pilot-specific activities throughout the project timeline.
5. Assessment of improvement progress, conducted through structured interviews with pilot leaders to evaluate the current level of advancement. (last months of the project)

It is important to note that this methodology does not aim to directly evaluate KPIs; rather, it focuses on assessing the overall production performance of the company in the context of its digital transformation journey.

4.2 Analysis of the Performance Pillar (AS-IS) – Survey

This section presents an analysis of the first iteration of survey responses (Annex 1), The primary objective of this survey is to capture the initial state (As-Is) of digital and organizational maturity within the manufacturing enterprise, while also identifying their future expectations (expected To-Be) and targeted development goals.

Following sections represent collected responses from the two main pilots under WP5 Avio and GF. For the remaining two pilots, AVL and VW, relevant information can be found in Deliverable D4.3.

As outlined earlier, the analysis is divided into two main components. The first focuses on the six core aspects of the Performance dimension, offering an interpretation of responses related to each area. The second addresses the additional questions that were specifically developed to align with the project and work package requirements, along with their corresponding analysis and interpretation.

4.2.1 Integrated Machine Tool Performance Self Optimisation Pilot (GF)-AS-IS

As depicted by Table 14, the partner had been provided "N/A" for four of the six categories: Operational-Technical, Economic, Environmental, and Social, leaving both AS-IS and TO-BE levels marked as Not Available. This indicates that, at the initial stage, these areas either have not been evaluated internally or are not yet in scope for structured measurement within the organization.

For the Product-Service Lifecycle, the partner reported a current maturity level of "Integrated", meaning that they already include Life Cycle Costing (LCC) and Environmental Life Cycle Assessment (LCA) in their analysis, aligning with circular economy principles. Looking ahead, their goal is to reach the "Exploited" level, which involves extending the assessment further to include Social LCA, contributing to broader sustainability and Green Deal objectives.

In terms of Supply Chain performance, the partner is currently at the "Integrated" level and intends to maintain this position moving forward. This reflects an established capability in measuring physical, economic, and sustainability-related KPIs, suggesting that the company already has a comprehensive approach in place for evaluating supply chain activities. The absence of change in the TO-BE status may indicate that current practices are seen as adequate or already aligned with strategic goals.

Table 14 – Summary of results of 1st Iteration GF

Performance Dimension	AS-IS	TO-BE
Operational - Technical	N/A	N/A
Economic	N/A	N/A
Environmental	N/A	N/A

Social	N/A	N/A
Product-Service Lifecycle	Integrated	Exploited
Supply Chain	Integrated	Integrated

Pilot business processes: The partner rated the progress in achieving objectives for process planning and preparation as "Good", and additionally, among the key milestones achieved, the partner highlighted the integration of Toolexpert into the CAD/CAM environment and the implementation of a Virtual Environment, marking important steps toward digitalizing and optimizing process planning workflows. However, a significant challenge encountered was the full implementation of data exchange across all partners, which remains difficult due to the complexity and variability of the environments involved.

Data Sharing and Integration: The partner rated the progress in connecting data sources and synchronizing data at the pilot site as "Good progress", reflecting a positive progress in the implementation of data integration activities. A challenge noted in this process was the presence of issues with data synchronization, indicating that while connectivity is being established, maintaining consistent and aligned data flows remains an area for improvement. Despite this, the partner evaluated the Data Container as “Very effective” in enabling data exchange and service implementation. A key benefit identified was the ability to implement new digital services, demonstrating how the Data Container is supporting the partner’s broader digital transformation objectives.

AI Models and Federated Learning: The partner reported "Good progress" in leveraging Federated Learning at the pilot site, indicating that the initial implementation is moving forward as planned. However, they rated the effectiveness of Federated Learning in enhancing AI models as "Somewhat effective”. Additionally, a key benefit identified was enhanced data privacy, reflecting the strength of Federated Learning in enabling distributed AI without centralizing sensitive information. However, the partner also noted a key challenge in the integration of Federated Learning with existing systems, which may require additional alignment of infrastructure and processes.

Progress in Integration: The partner rated the progress in integrating the RE4DY components and achieving the planned objectives at the pilot site as "Good", indicating steady progress in aligning system components and digital tools with project goals. Among the key achievements, the partner reported the successful connection of data sources to the Data Container and the completion of significant data synchronization tasks, both of which are essential steps toward enabling seamless data flow and interoperability across systems. However, the partner also identified technical difficulties with data integration as a major challenge, suggesting that despite progress, further work is needed to ensure reliable and efficient integration of components within the existing infrastructure.

4.2.2 Multi-Plant Predictive ZDM Turbine Production Pilot (AVIO)-AS-IS

Responses on performance’s aspects, reported in Table 15 and Figure 85 , revealed a clear understanding of diagnostic practices in several areas, while also highlighting the ambition to advance toward predictive and prescriptive capabilities supported by digital tools and AI. In terms of Operational and Technical performance, the company currently operates at a Defined level, meaning KPIs are well-structured, and efforts are made to understand the causes of performance trends. However, the expected target is to reach the Exploited stage, where AI/ML models and optimization tools enable simulation-based decision-making. Similarly, for Economic performance, the enterprise seeks to evolve from its current Defined stage, focused on causal analysis to an integrated level, where predictive models and financial forecasts guide proactive strategies. The Environmental and Product-Service Lifecycle dimensions are currently positioned at a Managed level, with efforts directed toward increasing structure and diagnostic capability (Defined). Regarding Social performance, the company reports a stable Defined level both in the current and future state, suggesting that existing methods for monitoring welfare and social indicators are satisfactory. Finally, Supply Chain performance is also currently at a Defined level, where both physical and economic indicators are measured. The desired shift to the Integrated stage reflects an intention to incorporate sustainability metrics into the evaluation framework, enabling a broader and more responsible performance assessment.

Table 15 – Summary of results of 1st Iteration Avio

Performance Dimension	AS-IS	TO-BE
Operational - Technical	Defined	Exploited
Economic	Defined	Integrated
Environmental	Managed	Defined
Social	Defined	Defined
Product-Service Lifecycle	Managed	Defined
Supply Chain	Defined	Integrated

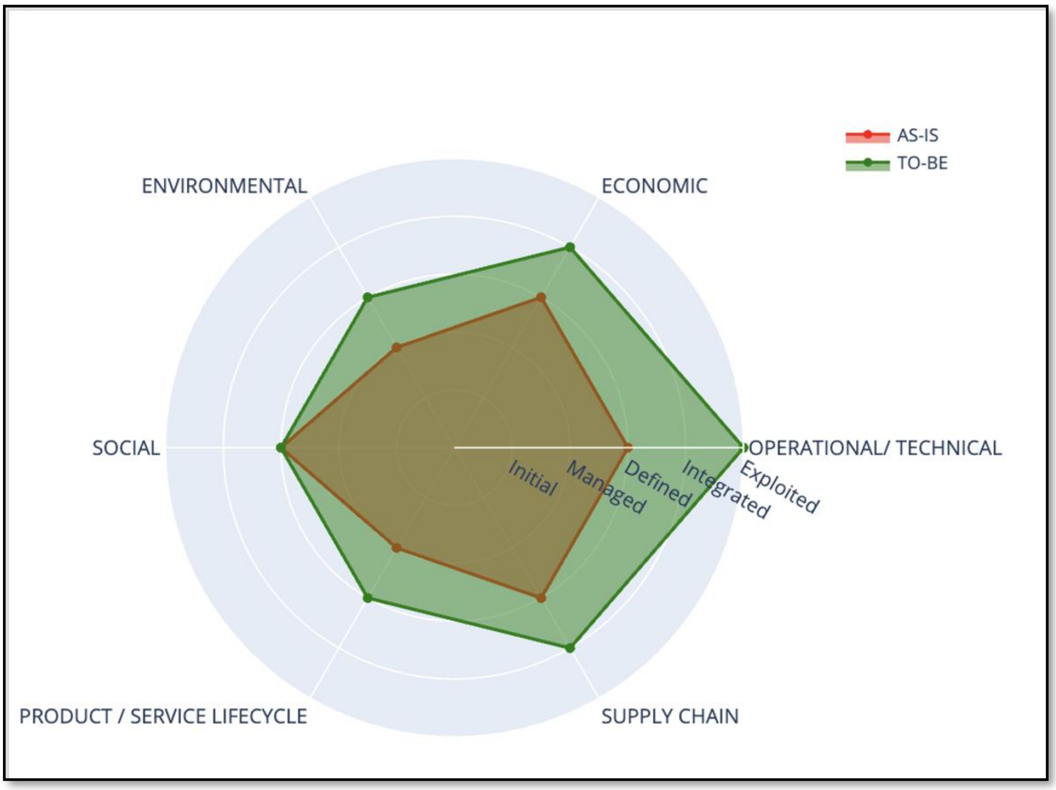


Figure 85 – Radar chart – Performance Avio

Pilot business processes: The partner rated the current status as “Good”, indicating a solid progress in aligning operational activities with the expected project outcomes.

In terms of key milestones achieved, the partner reported several notable accomplishments. These include the establishment of a shared set of business requirements, the development and execution of trial solutions leveraging the Testing and Experimentation Facility (TEF), and the definition of the architectural design for the intended solution. Additionally, an important operational step, an on-premises data acquisition campaign was successfully completed, providing the foundational data required for further implementation and integration efforts. Despite these advancements, the partner also highlighted a number of challenges and bottlenecks encountered during implementation. Two main issues were noted: first, the cybersecurity requirements associated with deploying the solution in an on-premise environment, particularly in the context of scaling up; and second, the difficulty of fitting the reference architecture into a real industrial setting while ensuring full compliance with industry-grade standards. These challenges underscore the complexity of translating architectural frameworks into practical, secure, and scalable industrial applications.

Data Sharing and Integration: The partner rated the progress in connecting data sources and synchronizing data at the pilot site as "Good progress". However, the main challenge encountered during the data sharing and integration process was identified as difficulty connecting data sources. This indicates that while the overall progress is positive, technical barriers still need to be addressed to achieve seamless connectivity across all systems. Regarding the effectiveness of the Data Container in enabling data exchange and service implementation, initially the partner indicated it was "Not very effective" while in the next phases, the Avio Aero-specific implementation of Data Container enabled interoperability between the two data formats at data source (edge) that required by the analytics components. However, a key benefit observed from using the Data Container was improved data visibility and transparency.

AI Models and Federated Learning: The partner rated the progress in leveraging Federated Learning at the pilot site as “Good progress”. However, they assessed its effectiveness in enhancing AI models as “Neutral”, suggesting that tangible improvements in model performance have yet to be realized. Consistent with this, the partner reported “No significant benefits observed” so far from implementing Federated Learning. Interestingly, despite the limited perceived impact, the partner indicated “No challenges faced”. It was further noted that any technical aspects or limitations should be addressed in coordination with the partners directly involved in the development of the AI models.

Progress in Integration: The partner rated the progress in integrating the RE4DY components and achieving the planned objectives at the pilot site as "Good". However, none of the listed key achievements, including data container connection, predictive maintenance implementation, digital twin development, or data synchronization have been fully reached at the current stage. Instead, the partner clarified that the team is currently working on deploying all these solutions into the production environment, indicating that implementation is still in progress. The main challenge encountered during integration has been related to technical difficulties with data integration. Additionally, the partner noted that the challenges reported in this section are consistent with those mentioned earlier in pilot business process section.

4.3 Analysis of the Performance Pillar (TO-BE) – Interview

As part of the final evaluation phase, a follow-up interview was conducted with the pilot representative to assess the latest progress in performance dimension, and to reflect on the achievements and challenges reported in the initial survey. Below are the responses obtained during the interview, which will be presented and analyzed on a section-by-section basis.

In addition, a structured interview was conducted to assess progress on KPIs (see Annex 2) beyond the factory-level pilot. The questions focused on performance trends, operational changes, data quality, operator feedback, challenges & mitigations, and future priorities, in line with the objective of generalizing impact insights at the pilot area level. This qualitative approach complements the quantitative KPI data collected during the project and supports broader communication and replication goals across the RE4DY value network. The interview content directly aligns with the expectations outlined in the task description, which emphasizes monitoring and evaluating transformed and blended data outputs, and extracting generalizable insights. By focusing on performance evolution since M24, data reliability, operational feedback, and lessons learned, the interview offers a comprehensive view of the pilot’s operational performance.

4.3.1 Integrated Machine Tool Performance Self-Optimization Pilot (GF) – TO-BE

Pilot business processes

- Progress Update on Process Planning and Implementation

The partner confirmed that the progress remains at a "Good" level, as previously reported. They emphasized that a complete set of application developments has been implemented, each corresponding to the four business processes defined in the project deliverables. Despite experiencing some delays in implementing the last set of business processes, the partner confirmed that these components are now ready and in place. Key milestones achieved include the successful development and deployment of applications covering process planning, preparation, and monitoring, specifically those leveraging AI. The current focus, as the project enters its final phase, is on testing all integrated components and preparing for the final update of KPIs, in alignment with the original objectives outlined in the trial handbook.

Data Sharing and Integration

- Data Container Implementation: Progress, Challenges, and Future Directions

In the interview, the partner reaffirmed their earlier assessment of "Good progress" in data integration, particularly in connecting and synchronizing data with the Data Container. When asked about the main technical and organizational challenges, the partner emphasized the complexity of collecting data from a diverse range of sources, including machines, tools, PLM systems, and software applications. Despite these challenges, they successfully consolidated this data, enabling enhanced collaboration among pilot participants such as Siemens, Innovalia, Metrology, Unimetric, GF, ATLANTIS, and others. Regarding the "data as a product" concept and its relevance for data sharing in a marketplace, the partner acknowledged the potential for sharing structured data externally, but noted that this would require well-defined agreements among stakeholders. Internally, agreements are already in place, making the concept feasible. However, broader external sharing through a data marketplace (such as the one proposed by ATOS) would need further discussion and clarification on legal and commercial frameworks. The partner described their experience with the Data Container as very effective, highlighting its critical role in enabling data synchronization and digital service implementation. They noted that without the container, these services could not have been deployed. The unified API and access control features were considered highly convenient, though the most critical and challenging aspect remained the synchronization and structuring of data from different systems. They recommended that future improvements could focus on standardizing data synchronization and structure, which they linked to concepts like Digital Product Passports. Enhancements in this area would help scale the Data Container's use across pilots and other industrial contexts.

AI Models and Federated Learning

- Federated Learning Deployment: Progress, Integration, and Predictive Use Case

During the interview, the partner provided additional insight into the implementation and effectiveness of Federated Learning in their pilot. They confirmed their previous evaluation of "Good progress", and noted that enhanced data privacy was a key benefit. However, they also reiterated that integration with existing systems posed a notable challenge. The partner explained that AI models were developed and trained using the RE4DY framework in close collaboration with CORE and ATLANTIS, who supported the implementation of Federated Learning architecture. One of the key challenges was harmonizing different technical requirements from multiple partners. Through joint efforts, a standardized and integrated Federated Learning solution was achieved, which the partner identified as one of the key deliverables of the project. Although no formal benchmarking data was yet available, the partner expects that aggregating data across systems in a federated manner will enhance model performance and scalability. The primary business use case where Federated Learning was applied focused on the prediction of tool lifetime in machine operations, marking a significant application of AI within the production process.

Progress in Integration

- RE4DY Framework Integration: Technical Achievements and Pathways to Standardization

In the interview, the partner confirmed the previously reported "Good progress" in integrating RE4DY components, including the Data Container, Federated Learning, and the Data-as-a-Product concept. They emphasized that these elements were all successfully implemented in the pilot and played a crucial role in enabling the technical achievements of the use case.

The partner explained that integration across layers of the RE4DY reference architecture was made possible through strong collaboration with partners such as CORE, ATLANTIS, and Siemens. However, one key challenge was the lack of standardization among systems and controllers used by different partners. Extracting consistent information from diverse machine controllers (e.g., Siemens and others) required additional effort to harmonize data content, not just format. The partner identified the standard definition of data structures, such as a Digital Product Passport, as a major learning outcome and an essential step for future scalability. They also highlighted the potential value of including mapping tools in the RE4DY architecture to automate the conversion of data from various controller formats into a standardized structure suitable for the Data Container or data marketplace. From a business perspective, the partner noted that the objectives of the pilot could not have been achieved without the RE4DY framework. Access to harmonized, cross-partner data enabled the delivery of accurate, valuable KPI predictions, which will bring long-term benefits as these solutions are further industrialized and brought to market.

Comparison with Initial AS-IS Results

In the initial analysis, the partner assessed their maturity as "Integrated" for both the Product-Service Lifecycle and Supply Chain dimensions. The ambition for the Product-Service Lifecycle was to reach the "Exploited" level, expanding from economic and environmental indicators (LCC and LCA) toward incorporating Social LCA and more holistic sustainability metrics. In the final interview, the partner reaffirmed that data integration and synchronization across multiple systems (e.g., Siemens, Metrology, Innovalia) had been achieved through collaboration and structured data exchange, particularly within the data container and federated learning frameworks. These integrations laid the foundation for enhanced product and service traceability and predictive performance monitoring, aligning with the target of reaching an "Exploited" level in the Product-Service Lifecycle. While full implementation of Social LCA may still be evolving, the foundational steps to enable it, such as harmonized data structures and cross-platform interoperability, have been put in place. For the Supply Chain dimension, the partner maintained its "Integrated" level. However, the interview suggests consolidation and strengthening of this status. This includes a growing ability to track lifecycle and operational metrics across supply chain actors, which reinforces their readiness for broader ecosystem-level integration.

4.3.2 Integrated Machine Tool Performance Self-Optimization Pilot (GF)-KPI Discussion

Performance Overview & KPI Progress

The project status is described as good. GF planned a set of applications aligned to four business processes and, despite delays, the last applications have now been implemented, are being rolled out, and are entering a phase where all components are tested. The focus for the closing months is to update the KPIs based on these integrated applications. The team cannot yet provide precise uplift figures relative to Month-24 because the past year was devoted to completing the data-collection campaign and advancing the implementation to something market-ready; KPI re-evaluation is therefore scheduled for the end of the project, which is one reason a project extension was requested. Within the business scope, a central use case is prediction of tool lifetime for machines, but quantitative effects are not yet reported.

Operational Insights

Over the last year the most relevant operational change was enabling data access for two distinct families of CNC controllers. This decision broadened coverage to what the team believes is roughly eighty percent of the market but required duplicating integration efforts, which had not been anticipated at the outset. There were no specific incident-level disruptions affecting performance beyond project delays; GF notes organisational restructuring that slowed implementation and, in turn, delayed KPI updates and motivated the extension request.

Data Collection & Accuracy

The consortium invested early effort in a common data model, and GF reports that this upfront alignment avoided surprises in collection and limited data-quality issues. Data collection revolves around machine data made available through an edge-and-cloud setup. GF highlights an effective edge computer and an Azure-based toolchain, with applications packaged as Docker containers to structure the architecture and deployment. This combination gave the team flexibility to operate at edge and cloud levels as needed.

Operators' Feedback

Operator engagement was most intense during the training phase for the machine-learning components. Operators' work practices, such as leaving machines running overnight or through weekends, required the project to adjust data-collection procedures and labelling so that training data remained accurate and usable. Without this back-and-forth, the training phase would not have been feasible.

Challenges and Mitigation

The multi-source synchronisation problem was the dominant technical challenge, compounded by heterogeneity across controller vendors and partner systems. GF and partners mitigated this by building the data container–based integration, establishing the edge–cloud setup, and standardising enough to proceed with training and application rollout. Organisationally, GF's internal restructuring introduced delays that slowed implementation and postponed KPI updates; the project extension provided time to finish integration and shift attention back to KPI evaluation.

Lessons Learned

The pilot's ambition, four business processes with many partners, made a single large team unwieldy. Splitting into focused groups per business process, with coordination across groups, proved more efficient. The experience reinforces the need for standard data definitions that support digital product-passport use cases, as well as practical tooling for mapping and normalising controller outputs into the container schema.

Next Steps and Recommendations

Immediate priorities are to complete testing, release the ready applications, finalise agreements with partners, and then update the KPIs on the basis of the integrated solution. From a sustainability perspective, GF would like to keep the pilot group or an equivalent collaboration channel active after project close so that partners can access tools, services, and contacts as needed. In practical terms, the recommendations are to maintain the edge–cloud plus containerised-apps deployment pattern, pursue standardisation for synchronisation and digital product-passport-aligned data structures, introduce mapping tools for controller heterogeneity, and proceed with federated learning for the tool-lifetime use case while planning quantitative comparisons as KPI updates are executed.

4.3.3 Multi-Plant Predictive ZDM Turbine Production Pilot (AVIO)-ToBe

Pilot business processes

- Updates on Milestones and Emerging Challenges Since the Initial Survey

The partner confirmed that the progress remains positive and highlighted an important new milestone not captured in the initial survey: the successful deployment of the Alida solution in Avio's Virtual Private Cloud (VPC) with valuable collaboration and support of ENG. This step marked a significant evolution from the earlier architectural design phase to an actual operational implementation within the company's environment. The partner emphasized that deploying the solution within their internal infrastructure, rather than relying on an external setup was a non-trivial task and required considerable effort and coordination.

- Overcoming Bottlenecks in Integrating the Reference Architecture into the Production Environment

The partner explained that the initial plan to use an external infrastructure had to be revised due to cybersecurity constraints and internal data-sharing policies. As a result, the team had to reallocate internal resources and engage additional departments that were not originally part of the project team. On the technical side, ENG played a critical role in co-designing and adapting components of the architecture to make them compatible with Avio's internal systems. This collaborative effort enabled successful deployment in a real-world, constraint-heavy industrial environment, which the partner described as a major added value for the project. The ability to implement and scale the pilot solution in such a setting demonstrates the project's effectiveness and real-world applicability.

Data Sharing and Integration

- Challenges in Connecting Data Sources and Managing Production Data Volume

The partner confirmed that progress remained "Good", especially considering the complexity of their environment and the available resources. They noted that finalizing the connection of machines to the infrastructure was particularly challenging, and despite partial success, there is still a lack of full automation in the data pipeline. They emphasized that for some pilots, data was collected directly from machines, while for others, data such as defect images was uploaded manually for offline analysis. A key challenge highlighted was the absence of a digital thread, that is, a seamless, connected flow of information linking data points across systems. The partner explained that while they had data in various silos, they lacked the relational structure to connect quality notifications to specific manufacturing steps or machine data. This gap significantly limits their ability to trace defects or correlate insights across processes, underscoring the need for stronger data governance and integration frameworks.

- Relevance and Applicability of the 'Data as a Product' Concept within the Pilot Context

The partner noted that while they did not have the opportunity to fully explore this concept during the project, they recognized its potential value, particularly in contexts where data exchange with machine OEMs could contribute to collective learning and defect detection improvements. They also emphasized that even within their own organization, across geographically distributed plants, data sharing could offer internal value. However, legal, IP, and export control constraints still pose challenges for such exchanges. When asked about their experience with the Data Container tool provided by UPV, the partner reiterated that while its effectiveness for their specific case was limited, it did contribute to improved data visibility and transparency. They suggested that greater internal data maturity and stronger data governance practices are prerequisites to fully benefiting from such tools. The partner emphasized that their current data infrastructure lacks the foundational layers, such as defined ontologies, metadata, and structured relationships necessary to take full advantage of unified APIs and federated data platforms. To illustrate this point, the partner provided a practical example: although they could collect time-series data from a machine, it was still difficult to associate that data with a specific part's serial number or correlate it with quality notifications. This lack of structured metadata and integrated processes significantly hampers efforts to build a robust digital thread or adopt tools like the Data Container effectively.

AI Models and Federated Learning

- Updates on the Effectiveness of Federated Learning: Benefits and Challenges Since the Initial Survey

The partner explained that the neutral rating was largely influenced by data quality issues. The pilot operates in a high-value, low-volume industrial setting, which means that the datasets available, particularly for defect detection were small, sparse, and highly variable. This made it difficult for AI models, including those trained using Federated Learning, to generate accurate and confident results. The partner emphasized that the low maturity and inconsistent nature of the available data significantly limited the ability to demonstrate the added value of the approach. He further noted that although Federated Learning was used in a third business scenario involving data exchange between machines, the challenge remained the same, insufficient data structure and consistency. The lack of meaningful results in this scenario, however, was not due to limitations in the Federated Learning methodology itself, but rather the immature data infrastructure supporting the pilot. The partner acknowledged the potential of Federated Learning and viewed the implementation experience as valuable and relevant, despite the technical difficulties and limitations encountered in practice.

Progress in Integration

- Assessment of the Suitability of the RE4DY Reference Architecture for the Pilot Use Case: Suggestions for Adaptation and Improvement

The partner confirmed that the reference architecture was both sufficient and flexible for their needs. They indicated that no major elements were missing, and that the architecture provided a solid foundation for implementing their solution. While some components had to be adapted to fit their internal technological stack and organizational policies, particularly due to cloud constraints and cybersecurity requirements. These adaptations were expected and manageable within the framework's flexible design.

The partner emphasized that the RE4DY architecture functioned as a guiding structure, which could be customized to align with company-specific needs. For example, certain identity management or cybersecurity tools included in the reference framework could not be used as-is due to internal compliance policies, but this was not seen as a limitation of the architecture itself. Instead, the partner saw this adaptability as a strength, confirming that the architecture was efficient and sufficient to meet their pilot's technical and business requirements.

- Comparison with Initial AS-IS Results

The analysis of the Performance Dimension, based on both survey responses and interview insights, highlights significant progress made across key areas, while also identifying opportunities for continued development. In Operational and Technical performance, the pilot specified their initial level as Defined, moving toward the Exploited level, where AI/ML models and simulation tools are applied, is underway, though further work is needed to enhance data automation and establish a stronger digital thread. In the Economic dimension, performance is being effectively monitored, and the ambition to implement predictive financial models is clear. Achieving this will require improvements in data quality and volume, which the pilot is actively addressing. For Environmental and Product-Service Lifecycle areas, efforts to move from Managed to Defined show good momentum. While challenges such as data fragmentation persist, the groundwork has been laid, making this transition realistic with targeted enhancements. The Social performance area remains stable at a Defined level, with current tools deemed sufficient for ongoing monitoring. In the Supply Chain dimension, the pilot is progressing well and aiming to integrate sustainability metrics into existing evaluations. To fully reach the Integrated level, the focus will be on strengthening data governance and enabling secure, compliant data sharing across internal sites.

4.3.4 Multi-Plant Predictive ZDM Turbine Production Pilot (AVIO)-KPI Discussion

Performance Overview & KPI Progress

After the month-18 review, AVIO narrowed and clarified the KPI set. For the predictive-quality pilot on machines, the KPI was defined as OEE, with a forecast of a one-percentage-point increase at machine level. This remains a relevant and conservative expectation, but any observed OEE movement is difficult to attribute solely to the pilot because shop-floor teams are continuously improving processes in parallel. Moreover, the team did not established a clear correlation between machine signals and quality outcomes, so even a positive OEE trend may be driven by other activities. For the image/defect-recognition pilots, the KPI comparing algorithm accuracy to operators has not yet reached parity. The main reason is data quality: there are few defective examples per class and a large variety of defect types, so the models have insufficient, imbalanced training data. A separate KPI anticipated about a 10% reduction in training time via the inspection/training interface; user interest and early feedback are positive, but formal reductions cannot be claimed because regulated training hours remain fixed.

Operational Insights

Over the last 12 months, ongoing shop-floor optimizations unrelated to the pilot likely influenced KPIs, which complicates causal attribution for the OEE metric. A major operational milestone was deployment of the solution inside AVIO's environment: a chunk of the Alida stack was installed with ENGINEERING's support and component adaptations. This shift was driven by cybersecurity and data-sharing constraints that prevented using an external platform and required reallocation of resources and engagement of additional internal teams.

Data Collection & Accuracy

Data capture has progressed but still faces last-mile issues. One pilot streamed machine data, while the imaging pilot relied on capturing pictures with different systems and uploading them for offline analysis, so end-to-end automation is not yet in place. The most significant accuracy limitation is a gap in the digital thread: it is hard to link machine time series to the exact part serial numbers and to the specific operation and quality notifications, which leaves datasets siloed and weakly cross-referenced. These gaps are rooted in data-governance maturity, including ontology and cataloguing. Regarding tools and systems, AVIO can provide a precise list by reconciling the reference architecture with in-house systems together with the technical lead.

Operators' Feedback

Operators and manufacturing engineers responded very positively to the interface for training and inspection support. AVIO is preparing the final shop-floor sessions, and capturing operator comments in the final deliverable was suggested because the solution is appreciated even if daily use is not yet established. While the interface likely improves robustness and knowledge transfer and could reduce practical learning time, regulatory requirements mean mandated training hours cannot be reduced.

Challenges & Mitigation

Security and data-sharing constraints meant an external platform could not be used; the team mitigated this by deploying inside the VPC, engaging additional internal teams, and adapting components with the technology partner's support. Last-mile connectivity and automation of machine-to-platform data flows remain challenging. Digital-thread gaps, especially linking time series to serial numbers, operations, and quality notifications, limit analytical power. For vision use cases, limited and imbalanced defect data constrained model accuracy.

Lessons Learned

Future pilots should prefer simpler, better-instrumented use cases with more examples per class and fewer defect types to accelerate learning and measurable KPI lift. They should invest earlier in data governance and management, ontology, catalogues, ownership, so analytics and any data-container approach rest on solid foundations. They should also anticipate industrial constraints from day one and plan for on-prem/VPC deployment when external platforms are unlikely to be permitted.

Next Steps & Recommendations

The immediate priorities are to close the digital thread across MES, quality, and machine data; to formalize a data-governance program; and to improve model readiness by enriching the dataset, including exploring synthetic data where appropriate. Given the successful internal deployment, continuing with a secure-by-design, AVIO-hosted pattern will reduce future integration friction. Post-project, AVIO does not expect ongoing support from the consortium; occasional assistance from the technology partner for running the internal solution may be requested, but no standing consortium resources are foreseen.

5 Conclusions

The RE4DY industrial pilots are taking an important step toward manufacturing ecosystems which have been digitally transformed and powered by AI. The pilots show how to employ the RE4DY reference framework to integrate cutting-edge digital tools, federated AI models, and advanced data architectures from the initial stages in several business and industrial environments. among the most important technological advances are

- The ability to effectively incorporate vendor-specific tool databases and virtual commissioning processes into CAM software
- Federated predictive maintenance systems that respect data privacy while offering useful insights
- In-process metrology solutions which make adaptive manufacturing more intuitive
- AI-driven defect detection with explainability that boosts both operational quality and operator training

These implementations result in broad operational benefits, like reduced setup and inspection times, extended tool lifetime, higher machine uptime, lower maintenance costs, lower scrap rates, and higher production throughput. The pilots' KPI assessment illustrates tangible business impacts: GF Fraisa expects up to 30% tooling cost reductions and 10% carbon footprint decreases, alongside a potential increase in machine availability up to 95%. Avio Aero reports a 44% reduction in quality control time, significant enhancements in defect detection accuracy, a 25% reduction in training hours, and positive OEE trends on EDM machines. These results underscore the practical value of federated learning and AI within industrial environments, while simultaneously identifying challenges related to dataset quality, system integration, and regulatory compliance that require continued attention. The 6P performance methodology proves effective for monitoring digital maturity and guiding transformational strategies. While current maturity ranges from managed to integrated across dimensions such as Operational-Technical, Economic, Environmental, and Supply Chain, the outlook targets full exploitation where AI-powered decision making, lifecycle sustainability assessments, and seamless data interoperability are embedded in daily operations. This requires further standardization efforts, enhanced data synchronization, stronger digital threads, and scalable edge-cloud federated solutions.

Looking ahead, the RE4DY pilots lay a robust foundation for extending AI-driven digital transformation across broader industrial sectors and supply chains. Future initiatives should focus on scaling federated learning infrastructures, enriching and diversifying datasets, advancing annotation and data governance practices, and driving regulatory acceptance of AI-assisted inspection and training processes. Continued partnership and knowledge sharing among technological, industrial, and academic stakeholders will accelerate the realization of smarter, more sustainable, and highly competitive manufacturing enterprises in the digital age.

6 Annex 1

1st Iteration of analysis – (AS-IS situation)[6]

Dear Partners,

The 6Ps methodology is a comprehensive tool designed to aid enterprises in their digital transformation journey by thoroughly analyzing six key dimensions: product, process, platform, people, partnership, and performance. This methodology emphasizes the importance of enhancing both technical and socio-business aspects to achieve successful digital transformation.

For our survey, we are focusing solely on the **Performance** dimension. This pilot experiment will compare the initial and final performance levels to measure the impact on the company's production process.

The survey includes a series of multiple-choice questions specifically tailored to assess the Performance dimension.

As the project approaches its conclusion, participants will need to indicate their initial status before the project (**As-Is**) and the actual situation (**To-Be**).

The full compilation of the survey will take you approximately 15 minutes overall, but you're allowed to save your partial compilation and reprehend it after a while.

[6] Survey link: https://polimi.eu.qualtrics.com/jfe/form/SV_3JcrkAaAeymKx2S

The RE4DY project team will process the results of the survey only in order to draft a report. Your privacy, personal and company data protection will be guaranteed in conformity with the European Regulation (EU) 2016/679. Your data will be processed in a separate database from the results of the survey in order to guarantee the anonymity of the survey and will not link your data with other databases. For more information regarding the processing of your data, you can visit here.

- [ ] Agree

FIRST NAME

LAST NAME

E-MAIL

COMPANY

PERFORMANCE

Performance dimension aims at investigating what is the AS-IS status before the project and the desired level of control over your company's processes and activities.

- COMPILE THE PERFORMANCE SURVEY
- GO TO THE Conclusion

OPERATIONAL/ TECHNICAL

What approach does your company adopt for measuring operational performances (e.g. OEE)?

1. INITIAL: Operational performance is often not measured or understood
2. MANAGED: Descriptive Performance - Measurement and analysis of business KPIs are largely retrospective
3. DEFINED: Diagnostic Performance - Measurement of KPIs is clear. Attempt to understand the causes that affects events and behaviours
4. INTEGRATED: Predictive Performance - Measurement of KPIs is prospective. Statistical models are used to forecast and to understand the KPIs predictions
5. EXPLOITED: Prescriptive Performance – future-oriented. Optimization and simulation to find the best course of action and operational KPIs measurement. AI/ML models are used to forecast and to understand the KPIs predictions

	AS-IS	TO-BE
INITIAL	○	○
MANAGED	○	○
DEFINED	○	○
INTEGRATED	○	○
EXPLOITED	○	○
Not Applicable	○	○

ECONOMIC

What approach does your company adopt for measuring economic performances (e.g. ROI)?

1. INITIAL: Economic performance is often not measured or understood
2. MANAGED: Descriptive – Measurement of economic KPIs is largely retrospective
3. DEFINED: Diagnostic - Measurement of economic KPIs is clear. Attempt to understand the causes of events and behaviours
4. INTEGRATED: Predictive - Measurement of economic KPIs is prospective. Statistical models and forecasts techniques to understand the KPIs predictions
5. EXPLOITED: Prescriptive – future-oriented. An AI decision-making support system boosting optimization exploits simulation and allows to find the best course of actions and operational KPIs measurement

	AS-IS	TO-BE
INITIAL	◯	◯
MANAGED	◯	◯
DEFINED	◯	◯
INTEGRATED	◯	◯
EXPLOITED	◯	◯
Not Applicable	◯	◯

ENVIRONMENTAL

What approach does your company adopt for measuring environmental performances (e.g. water consumption per product, energy optimisation) ?

1. INITIAL: Environmental performance is often not measured or understood
2. MANAGED: Descriptive – Measurement of environmental KPIs is largely retrospective
3. DEFINED: Diagnostic - Measurement of environmental KPIs is clear. We attempt to understand the causes of events and behaviours
4. INTEGRATED: Predictive - Measurement of environmental KPIs is prospective. AI and/or statistical models are used to forecast environmental performances
5. EXPLOITED: Prescriptive – future-oriented. An AI decision-making support system boosting optimization exploits simulation and allows to find the best course of action and environmental KPIs measurement

	AS-IS	TO-BE
INITIAL	◯	◯
MANAGED	◯	◯
DEFINED	◯	◯
INTEGRATED	◯	◯
EXPLOITED	◯	◯
Not Applicable	◯	◯

SOCIAL

What approach does your company adopt for measuring social performances (e.g. welfare for employees)?

1. INITIAL: Social performance is often not measured or understood
2. MANAGED: Descriptive - Measurement of social KPIs is largely retrospective
3. DEFINED: Diagnostic - Measurement of social KPIs is clear. Attempt to understand the causes of events and behaviours
4. INTEGRATED: Predictive - Measurement of social KPIs is prospective. AI and/or statistical models are used to forecast social performances
5. EXPLOITED: Prescriptive – future-oriented. An AI decision-making support system boosting optimization exploits simulation and allows to find the best course of action and environmental KPIs measurement

	AS-IS	TO-BE
INITIAL	○	○
MANAGED	○	○
DEFINED	○	○
INTEGRATED	○	○
EXPLOITED	○	○
Not Applicable	○	○

PRODUCT-SERVICE LIFECYCLE

Which dimensions of analysis are taken into account in the assessment of lifecycle of the products/services offered to the customers?

1. INITIAL: No product life cycle assessment
2. MANAGED: A few life-cycle aspects are included in some KPIs, but occasionally
3. DEFINED: Life Cycle Costing (LCC) towards recycling, re-use, de- re-manufacturing KPIs
4. INTEGRATED: Life Cycle Costing + Environmental LCA towards Circular Economy
5. EXPLOITED: Life Cycle Costing + Environmental LCA + Social LCA towards Sustainability and Green Deal

	AS-IS	TO-BE
INITIAL	○	○
MANAGED	○	○
DEFINED	○	○
INTEGRATED	○	○
EXPLOITED	○	○
Not Applicable	○	○

SUPPLY CHAIN

Which dimensions of analysis are taken into account for the overall evaluation of your company's supply chain?

1. INITIAL: The Supply Chain performances are lowly monitored/measured.
2. MANAGED: We measure only the most important physical performance of suppliers (e.g. punctuality, quality, operational flexibility)
3. DEFINED: We measure physical and economical performances (purchase price, non-quality costs, delivery delays, lack of flexibility, etc.).
4. INTEGRATED: We measure physical and economical performances, and sustainability indexes.
5. EXPLOITED: We measure physical and economical performances, sustainability indexes and cross-company value creation.

	AS-IS	TO-BE
INITIAL	○	○
MANAGED	○	○
DEFINED	○	○
INTEGRATED	○	○
EXPLOITED	○	○
Not Applicable	○	○

Pilot business processes

How would you rate the progress in achieving objectives for process planning and preparation?

	Rate
Excellent	○
Good	○
Average	○
Poor	○
Insufficient	○

What key milestones have been reached in process planning and preparation? (Please list any significant achievements)

What challenges or bottlenecks have you encountered in implementing process planning and preparation? (Please specify)

Data Sharing and Integration

How would you rate the progress in connecting data sources and synchronizing data at your site on the dedicated Data Container?

	Rate
Excellent Progress	○
Good Progress	○
Average Progress	○
Poor Progress	○

What challenges have you faced in the data sharing and integration process?

	Challenges
Difficulty connecting data sources	○
Issues with data synchronization	○
Concerns about data security and privacy	○
Lack of resources or expertise	○
Other challenges (please specify)	○

OTHER CHALLENGES

How effective has the Data Container been in enabling data exchange and services implementation?

	Rate
Very effective	○
Somewhat effective	○
Neutral	○
Not very effective	○
Not at all effective	○

What benefits have you seen from using the Data Container for data sharing and integration?

	Benefits
Improved data visibility and transparency	○
Ability to implement new digital services	○
Enhanced data-driven decision making	○
Cost savings from streamlined data processes	○
Other benefits (please specify)	○

OTHER BENEFITS

AI Models and Federated Learning

How would you rate the progress in leveraging on Federated Learning at your site?

	Rate
Excellent progress	○
Good progress	○
Average progress	○
Poor progress	○
Insufficient progress	○

How effective has the Federated Learning approach been in enhancing AI models?

	Rate
Very effective	○
Somewhat effective	○
Neutral	○
Not very effective	○
Not at all effective	○

What benefits have you observed from implementing Federated Learning? (Select all that apply)

	Benefits
Improved model accuracy	☐
Enhanced data privacy	☐
Reduced data transfer costs	☐
No significant benefits observed	☐

What challenges have you faced in implementing Federated Learning? (Select all that apply)

	Challenges
Technical difficulties in deployment	☐
Data privacy concerns	☐
Lack of expertise	☐
Integration with existing systems	☐
No challenges faced	☐

Are there any other comments or feedback you would like to provide regarding AI models and Federated Learning in the RE4DY pilot?

Progress in Integration

How would you rate the progress in integrating the RE4DY components and achieving the planned objectives at your pilot site?

	Rate
Excellent	○
Good	○
Average	○
Poor	○
Insufficient	○

Which of the following key achievements have been reached at your pilot site? (Select all that apply)

	Achievements
Successful connection of data sources to the Data Container	☐
Implementation of predictive maintenance solutions	☐
Development of digital twins for tools and machines	☐
Completion of significant data synchronization tasks	☐
None of the above (please specify)	☐

None of the above (please specify)

What are the main challenges and roadblocks you have encountered in implementing the pilot at your site? (Select all that apply)

	Challenges
Technical difficulties with data integration	☐
Insufficient resources or expertise	☐
Resistance to change from staff	☐
Issues with data quality or availability	☐
No significant challenges faced	☐

Are there any other comments or feedback you would like to provide regarding the integration of RE4DY components at your pilot site?

In the next page, you can find a summary of your responses.

7 Annex 2

KPIs Interview Questions

Performance Overview & KPI Progress

- Which KPIs from D4.2 have improved since M24, and by how much? (E.g. percentage increase in yield, efficiency, uptime.)
- Are there any KPIs that have not met expected targets? What factors contributed to this?

Operational Insights

- What operational changes or process improvements have impacted KPIs during the last 12 months of the pilot?
- Can you highlight any incidents or disruptions that temporarily affected performance metrics?

Data Collection & Accuracy

- How reliable and timely has the data capture process been? Have there been any data gaps or quality issues?
- Which monitoring tools or systems provided the most valuable data for KPI monitoring?

Operators' Feedback

- What insights did you gather from operators (via surveys or interviews) that influenced KPI outcomes?
- Engagement with operators: how did their feedback shape operational adjustments?

Challenges & Mitigation

- What key challenges have you encountered since M24, and how have you addressed them?
- What lessons learned would you highlight to improve process operations for future pilots?

Next Steps & Recommendations

- Based on current KPI performance, what are your priorities going forward?
- Are there any support actions (resources, training, tools) required from the consortium to maintain or further improve performance?