D4.5 AI4Manufacturing Toolkit

Author:	CARTIF
Work Package:	WP4 Beyond PLATFORMS: AI DIH Open Platforms and DIH Platform
Delivery date:	15.10.2021
Due date:	30.09.2021
Classification:	Public
Type:	Other

H2020 Innovation Action - This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement N. 952003

Executive Summary

The AI REGIO project aims to lowering the barriers preventing AI-driven DIHs from implementing fully effective digital transformation pathways for their Manufacturing SMEs.

In this context, five subsystems are going to be designed and implemented in the context of WP4 "Beyond PLATFORMS: AI DIH Open Platforms and DIH platform" and WP5 "Beyond INDUSTRY 4.0: AI DIH Industry 5.0 and Data Sharing Spaces", which are.

1. Data4AI Platform, which is a new generation of platforms for ensuring data quality for AI applications based on syntaxis and semantic context.
2. AI4Manufacturing Toolkit, which is expected to be a set of data analytics tool and techniques where prepared and cleaned data can be exploited for enabling experimentation with AI technologies.
3. DIHIWARE Platform, which is an environment where providers and consumers of digital technologies related to AI development and adoption cannot just matching assets and needs, but they can collaborate to boost innovation.
4. Collaborative Intelligence Platform, which allows to design, monitor, optimize, and simulate the orchestration of human-centred processes workflow in terms of process management and Human-AI interaction.
5. AI Data Spaces for Manufacturing, which is data sovereignty solution based on Industrial Data Space (IDS) Open Source current initiatives.

WP4 is focused on 3 of these 5 subsystems: Data4AI (WP4.2), AI4Manufacturing (WP4.3) and DIHIWARE (WP4.4). Furthermore, WP4.3 "AI4Manufacturing Toolkit and Innovative AI Ecosystem" is focus to design and implement AI4Manufacturing Toolkit as a set of data analytics tools and techniques allowing the integration of different relevant outcomes in terms of algorithms, software frameworks, development tools and datasets.

This deliverable D4.5 is accompanying the first iteration of task WP4.3 hat has taken place until M12. In these first months of development, the tasks have focused on the analysis of the requirements and pre-selection of a set of Open Source and proprietary components that will make up the catalogue of the Toolkit.

Contents

Figures

Tables

Table of acronyms

AI	Artificial Intelligence
DIH	Digital Innovation Hub
DoA	Description of Action
IDS	Industrial Data Space
OSS	Open Source Software
SME	Small to Medium Enterprise

1 Introduction

1.1 Scope of the Deliverable

The main objective of WP4 is define a Reference Architecture for AI for manufacturing systems by aligning the state-of-the-art Industry 4.0 Reference Architectures and Reference Models from previous I4MS and AI-related initiatives. In that sense, D4.1 "Reference Architecture for AI DIH" provide the AI REGIO Reference Architecture, among others, for the AI4Manufacturing Toolkit subsystem to be developed within WP4.3. Furthermore, a Reference Implementation, based on Open Source components, is defined and implemented in order to lower the barriers for the adopters, acting as a baseline for the experiment deployments.

Specifically, the aim of WP4.3 is design and implement a set of data analytics tools and techniques, named AI4Manufacturing toolkit, using the main outcome of WP4.2 (named Data4AI platform), where prepared and cleaned data can be exploited for enabling experimentation with AI technologies. A special attention has been paid to new AI assets coming from SMEs, I4MS projects and AI4EU initiative.

1.2 Structure of the Document

The deliverable is structured as follows:

- Section 2 summarizes some foundational aspects related to the Toolkit: the user journey, the requirements, and the architecture overview. Moreover, a first Open Source based implementation is proposed.
- Section 3 is focused on the description of the main AI assets available in the environment of AI REGIO project to build the Toolkit.
- Section 4 details how to get access to the Toolkit.
- Finally, section 5 reports the conclusions and the steps to follow through second iteration of development.

2 AI4Manufacturing Toolkit

In this section some foundational aspects related to the AI4Manufacturing Toolkit are presented, as user journey, basic requirements, architecture description from a more technical point of view and first Open Source reference implementation.

2.1 User Journey and Grand Scenarios

The main potential users/adopters of the AI4manufacturing Toolkit identified up to now are the AI system developers/integrators. In particular, they are Data engineers, Data Scientists and/or System integrators developing AI-based applications useful for manufacturing domain, mainly SMEs. Through the user journey, they have been described in the following. Furthermore, the associated grand scenarios explaining their main role is reported.

Stakeholder Group Name	**AI system developers/integrators**
Tasks	Develop AI-based solution to be used within manufacturing environment
Personal interests	A technological provider that analyses, designs and writes code to build AI solutions, and develop manufacturing applications powered by AI models.
Required expertise (description and level)	A person skilled in AI with some basic knowledge in manufacturing domain and knowledge of AI tools interoperability mechanisms (e.g., API REST for information exchange)
Social Environment	Continuous collaboration with manufacturing SMEs workers
Physical Environment	Someone working in a technology company developing applications for manufacturing SMEs
Comments	

GS2.1	***Maria acting as an AI system developer who is looking for an AI technological asset within manufacturing domain and wants to know how to work with it***
Stakeholder(s)	AI system developers/integrators
Scenario	Once authenticated in AIREGIO portal, Maria is able to access to its catalogue, search for a specific asset and see if it is available in the AI4Manufacturing Toolkit. If available, she can: • Know more specific information about the asset • Obtain recommendations/guidelines about how to use a specific AI asset available in the Toolkit.

GS2.2	***Maria acting as an AI system developer who is looking for a designer environment where build AI applications within manufacturing domain***
Stakeholder(s)	AI system developers/integrators
Scenario	Once authenticated in AIREGIO portal, Maria is able to access to AI4Manufacturing Toolkit and: • Create AI Pipelines by composing AI Resources in a workflow. Each pipeline is modelled as a direct acyclic graph and can be executed in a specific production environment, compatible with the designer environment where they were created.

GS2.3	***Maria acting as an AI system developer who is looking for an AI REGIO template to build an AI application within manufacturing domain***
Stakeholder(s)	AI system developers/integrators
Scenario	Once authenticated in AIREGIO portal, Maria is able to access to AI4Manufacturing Toolkit and: • Look for a pre-defined AI REGIO template in some of the well-known designer AI pipelines environment supported by the Toolkit (e.g., predictive maintenance, defect detection, …). • Select the desired template and modify the AI pipeline to incorporate her own AI Resources and Datasets

2.2 Requirements

The Toolkit requirements, already presented in the D4.1 "Reference Architecture for AI DIH", can be summarized as:

- The Toolkit must fill the gap identified in the absence of AI resources for manufacturing environment, especially for SMEs. Moreover, the Toolkit must guide the user in the selection of existing well-stablish technologies.
- The Toolkit must be implemented avoiding "reinventing the wheel", so using existing Open Source Software (OSS) technologies. Also "reusing as much as possible from partners", so proprietary solution must also be analyzed in order to maximize the reuse of backgrounds and existing components as well as providing more flexibility in developing new AI assets.
- The Toolkit must be interoperable with other platform in the community (e.g., AI4EU) and well-known Open Source solutions, and fully functional even without the connection with external platforms. Moreover, developers and deployers should have "freedom of choice", so not limit them to a single platform but a collection on multiple ones.

2.3 Architecture Overview and Ecosystem

AI4Manufacturing Toolkit (Figure 1) can be seen as a collection of operational technologies, data analytics tools and platforms, designed to provide support to system integrators and technology adopters to create new AI-based applications. In a more formal way, the Toolkit can be composed by three main types of entities:

- **AI resources**, seen these as reusable machine learning or deep learning models to solve a specific manufacturing problem.
- **AI pipelines**, where a set of AI resources are composed in a workflow.
- **AI orchestrators**, as end-to-end application where several AI pipelines or AI resources are managed/ orchestrated.

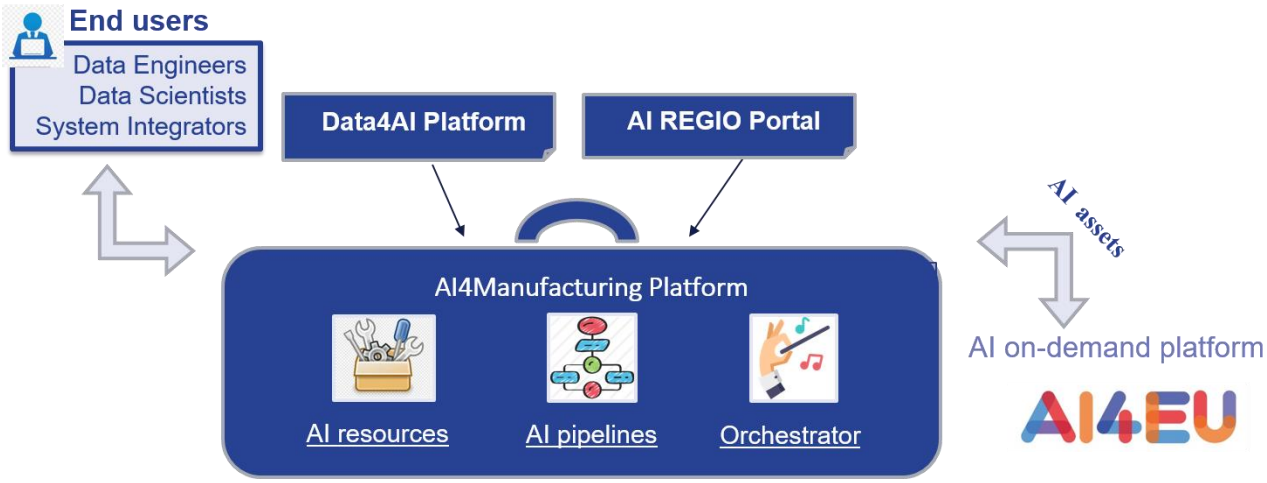

Figure 1 AI4manufacturing Architecture Description and Ecosystem

Of course, the Toolkit is not isolated in the ecosystem of AI REGIO project. The Toolkit is closely interrelated with the Data4AI Platform where data can be prepared and cleaned. Clarify here that, not only data coming from Data4AI can be exploited in AI4Manufcturing Toolkit, but also end users can experiment with their own datasets.

Also, with the AI REGIO Portal where the searchable catalogue of AI resource is available. Again, say that end users will access to AI resources and pipeline through this portal or test/experiment with their own ones.

Finally, as external stakeholders, highlight the European AI on-demand platform (AI4EU). In that sense our objective is to guarantee the interoperability between AI assets, so that the assets can be exchanged between both platforms.

2.4 Open Source Implementation Landscape

As requirements, a first Open Source based reference implementation (Figure 2), has already been presented in D4.1 "Reference Architecture for AI DIH". Main advantages of using Open Source elements are to avoid vendor lock-in, lower costs and barriers for the adopters, interoperability, and user adoption.

Figure 2 AI4Manufacturing Open Source Reference Implementation

Besides this Open Source implementation, also proprietary solutions have been taking into account to implement the Toolkit. In that way, next section is focus on describe and analyse technologies coming from the partners and already listed in the DoA (Annex B).

3 AI Assets: backgrounds/foreground

Beyond the Open Source reference implementation proposed previously, there are a set of 39 technological assets as initial content of the AI4Manufacturing Toolkit (Table 1). This list, collected at proposal time, have been analysed and matched to classify them between the three main types of entities defined. As can be seen in the table, not all the technological assets are useful for the AI4Manufacturing Toolkit[1], since they are useful for the other Platform to be generated/developed in AI REGIO project (i.e. Data4AI, Data Spaces for Manufacturing, Data Models and Ontologies etc..).

Technology Asset	Technology Provider	Type
A decision support system for troubleshooting	Intellimech	AI pipeline
ALIDA	ENG	AI orchestrator
Answer Set Programming (ASP)	STIIMA-CNR	*deprecated*
BELTZBOX	TECNALIA	*∉ toolkit*
Converting 3D-Scan into closed 3D-Model	FZI	AI pipeline
CPS Virtual factory	SouthNetherlandsDIH	AI orchestrator
CROMADOS	TECNALIA	*∉ toolkit*
D2 Anomaly Analyzer and Predictor	Nissatech	AI resource
D2Twin	Nissatech	AI orchestrator
Data diode	SouthNetherlandsDIH	*∉ toolkit*
Dezyne	SouthNetherlandsDIH	AI orchestrator
Digital Industry Data Analytics (DIDA) Platform	ENG	AI orchestrator
DIHIWARE	ENG	*deprecated*
Digital Virtualization Experience Suite (DIVE)	ENG	*deprecated*
Dynamic Planner & Scheduler	Eurecat	AI resource
Factory Data Model	STIIMA-CNR	*∉ toolkit*
InformationGrid	SouthNetherlandsDIH	*∉ toolkit*
Knowage	ENG	*∉ toolkit*
LEO+DASWIND	TECNALIA	*∉ toolkit*
Machine learning predictive maintenance	STIIMA-CNR	AI resource
Matchmaking methodology and tool	TAU	*∉ toolkit*
OPMOPRO	TECNALIA	AI resource
Pacelab WEAVR	TXT	AI orchestrator
Papyrus4Manufacturing	CEA	*∉ toolkit*
Process integrated feedback management	FZI	AI pipeline
Process monitoring and optimization toolkit (POCAS)	SKU	AI resource
Product, Manufacturing Resource and Capability Ontologies	TAU	*∉ toolkit*
REFILOC	TECNALIA	*∉ toolkit*
S5 Blockchain Data Brokerage Engine	Suite5	*∉ toolkit*
S5 Enterprise Big Data Analytics Suite-Manufacturing	Suite5	AI orchestrator
S5 Enterprise Big Data Analytics Suite-Social	Suite5	*∉ toolkit*

[1] *∉ toolkit*

Semantic Middleware Enabling the Factory Telemetry	STIIMA-CNR	*∉ toolkit*
Smart IoT Analytics Platform (SIoTAP) and Tools	UM	AI orchestrator
StreamPipes	FZI	*∉ toolkit*
Summar.io	SouthNetherlandsDIH	*∉ toolkit*
Supervised 2D based scene comparison system	Intellimech	AI resource
Supervised real-time 2D-based Object Detection system.	Intellimech	AI resource
SURFIN	TECNALIA	*∉ toolkit*
TXT IIOT Platform	TXT	*deprecated*

Table 1 Technological assets described in the DoA (Annex B)

Since AI4Manufacturing Toolkit will take into account, not only the AI assets already listed, but also all those new assets that can be generated in the experiments and the open calls, the previous list is dynamic so that its content will evolve continuously. Moreover, the categorization may also be modified as the AI asset improves.

Following sections show some examples of AI assets available in the AI4manufacturing Toolkit: background technologies, plans for the improvements, how to access it and a short factsheet.

3.1 AI Orchestrator: an example

As an AI orchestrator, the "S5 Enterprise Big Data Analytics Suite-Manufacturing" asset is a proprietary solution whose main aim is to offer descriptive and predictive analytics for manufacturing legacy and operational data related to maintenance and deliver insights on critical manufacturing operations (e.g., if there will be a breakdown for a machine the next day) in an end-to-end manner. Apart from the analytics and visualisation services, the engine features also data collection, mapping and pre-processing capabilities that are necessary to perform the rest of its core functionalities.

In the context of AI REGIO and in order to improve and adapt the asset for AI4Manufacturing Toolkit, the "S5 Enterprise Big Data Analytics Suite-Manufacturing" aims to enhance and improve its offerings, through the actual system validation in additional operational manufacturing environments. Additional AI-driven visualization dashboards and their associated ML/DL analytics are developed based on the analysis of the AI REGIO DIH and pilot cases. Notifications shall be provided for prediction and recommendation decision support, while the available algorithms and the pre-trained models will be extended, retrained as required and scheduled for execution as required. The dataset basis, that now hosts synthetic manufacturing datasets, will be expanded with additional datasets from the AI REGIO project. New data ingestion methods, such as the utilisation of IDS connectors are also explored to make the"S5 Enterprise Big Data Analytics Suite-Manufacturing" more flexible and interoperable with external systems. Finally, the underlying manufacturing data models will be appropriately extended – whenever needed - to cover the needs that will emerge from the experimentation and validation activities in AI REGIO. Table 2 shows the factsheet written for this AI Orchestrator.

S5 Enterprise Big Data Analytics Suite-Manufacturing			
	"Background"		**"Foreground"**
General description	WHAT? A data-at-rest analytics engine driven by the need to leverage manufacturers' legacy data and operational data related to maintenance, and to extract and correlate relevant knowledge HOW? Through an intuitive business user interface targeted at business users, based on the analysis configured by a data analyst (to a separate data analyst interface)	»	• Extending the pre-trained models supported in an out-of-the box manner to support the AI REGIO use cases and additional manufacturing problems • Supporting additional data sources • Increased data security/privacy • Supporting rule-based notifications (e.g., for predictions beyond a certain threshold) and scheduling of analytics execution
Use/Distribution	Closed Source/Proprietary solution	»	Closed Source/Proprietary solution

Key AI features	• Supported AI Models for: Time Series Prediction, Dimensionality Reduction, Clustering, Classification, Regression • Interactive data exploration • Results visualisation in dashboards • Data: synthetic manufacturing datasets • Individual, separate interfaces for different user groups (data scientists and business users)	»	• Support for additional AI models and AI-driven Applications • Additional visualisations for AI analytics results • Scheduling of the AI models execution • Rule-based notifications for AI execution progress • Additional datasets from the AI REGIO DIHs and pilot cases
Technology stack	Built on top of the following Open Source technologies: 	»	Additional AI libraries: Spark MLlib Additional Visualization Library: ApexCharts Orchestration: Kubernetes Packaging: Docker
Interoperability mechanisms	• Underlying data models for semantic interoperability • Handle csv files	»	• Extend the underlying manufacturing data models • Data exchange through IDS connectors • Handle additional file formats (e.g. JSON, XML, etc.)
Access	Confidential, Access available only to the subscribed clients	»	Access to online demo will be provided within the AI REGIO consortium to allow the DIHs and pilot cases perform their experiments
Validation	Utilized in predictive manufacturing business cases	»	*To be decided*
Owner	Suite5 Data Intelligence Solutions		

Table 2 AI Orchestrator factsheet "S5 Enterprise Big Data Analytics Suite-Manufacturing"

3.2 AI Pipeline: an example

AI REGIO will offer domain experts and manufacturing data scientists with access to Jupyter Notebooks for popular manufacturing use cases (i.e., predictive maintenance, RUL estimation, defect detection) to be used for:

- **Training and Education on industrial use cases**. Specifically, the Jupyter notebooks will provide practical examples that will accompany training presentations and webinars destined to SMEs, including trainings delivered in the scope of the Digital Innovation Hubs of the project.
- **Replication/Reuse following a fair effort for adaptation to different production lines**. The notebooks will serve as practical examples and demonstrators for manufacturing enterprises that are interested in deploying AI/ML in their production lines. The notebooks will bootstrap their development efforts through providing readily available examples that can be modified and customized to support similar manufacturing use cases. In this direction, relevant instructions will be provided.

Table 3 shows the factsheet generated for this AI Pipeline.

Jupyter Notebooks for popular manufacturing use cases			
	"Background"		**"Foreground"**
General description	Publicly Accessible Open Source Jupyter Notebooks on Machine Learning Techniques for Industrial Use Cases	»	AI REGIO Customized Notebooks addressing popular ML use cases for Manufacturing
Use/Distribution	General Education, Training and Awareness on Machine Learning, Deep Learning and Artificial Intelligence	»	Education, Training and Awareness on Machine Learning for Manufacturing Use Cases -Code Reuse
Key AI features	• Machine Learning (Regression) • Deep Learning (Long Short Term Memory) • Predictive Analytics on open source datasets	»	• Machine Learning (Regression) • Deep Learning (Long Short Term Memory) • Predictive Analytics on industrial/manufacturing datasets
Technology stack	jupyter TensorFlow K Keras	»	jupyter TensorFlow K Keras

Interoperability mechanisms	Python Libraries Compatibility Dataset standardization (CSV, XML, JSON)		Python Libraries Compatibility Dataset standardization (CSV, XML, JSON) Dataset preprocessing to follow standards based format
Access	Github		Github
Validation	N/A		N/A
Owner / Developer	ExpertAI offered with a Public / Open Source license such as GPL or MPL (for the software) and Creative Commons (for the documentation)		

Table 3 AI Pipeline factsheet "Jupyter Notebooks for popular manufacturing use cases"

3.3 AI Resource: an example

"AI-based PREDICTIVE DYNAMIC PRODUCTION PLANNER" asset main aim it to support production managers in the phase of acquisition of raw materials and organization of production. It includes a demand forecasting module, which is based on Machine Learning techniques including Extreme Gradient Boosting Regressor and work orders sequencer based on Metaheuristics. Both solutions take advantage of Open Source libraries such as scikit-learn and OptaPlanner. Table 4 shows the factsheet elaborated for this AI Resource.

Dynamic Planner & Scheduler			
	"Background"		**"Foreground"**
General description	Demand forecasting module, based on Extreme Gradient Boosting Regressor, fully adapted to the use case and exploited as library Work orders sequencer module fully adapted to the use case and exploited as library	»	• Generalizing the work orders sequencer to support multiple use cases with minor configurations • Enhancing the accuracy of demand predictions by using other technologies like LSTM neural networks • Supporting the interoperability of the asset by creating an API REST
Use/Distribution	Closed Source/Proprietary solution	»	Closed Source/Proprietary solution
Key AI features	<u>Demand forecasting module</u> Based on Ensembled Machine Learning techniques <u>Work Orders Sequencer</u> Based on metaheuristics	»	<u>Demand forecasting module</u> Based on Deep Learning techniques <u>Work Orders Sequencer</u> Based on metaheuristics

Technology stack	Demand forecasting module Programming language: python AI libraries: scikit learn Work Orders Sequencer Programming language: Java AI libraries: OptaPlanner Z3 Theorem Prover	⟫	Demand forecasting module Programming language: python AI libraries: PyTorch Deployment: Work Orders Sequencer Programming language: Java AI libraries: OptaPlanner Z3 Theorem Prover Deployment:
Interoperability mechanisms	Demand forecasting module Embedded directly on existing solutions without interoperability mechanisms, based on calls to the libraries Work Orders Sequencer Embedded directly on existing solutions without interoperability mechanisms, based on calls to the libraries	⟫	Demand forecasting module Based on REST API Work Orders Sequencer Based on REST API
Access	Demand forecasting module Access to code through internal Eurecat ICT infrastructures (internal github) Work Orders Sequencer	⟫	Demand forecasting module Access to code through internal Eurecat ICT infrastructures (internal github) Work Orders Sequencer

	Access to code through internal Eurecat ICT infrastructures (internal github)		Access to code through internal Eurecat ICT infrastructures (internal github)
Validation	<u>Demand Forecasting module</u> Validated at two levels, laboratory (cross-validation with historical data) and operational environment <u>Work Orders Sequencer</u> Validated at two levels, laboratory (expert knowledge) and operational environment		<u>Demand Forecasting module</u> Validated at two levels, laboratory (cross-validation with historical data) and operational environment <u>Work Orders Sequencer</u> Validated at two levels, laboratory (expert knowledge) and operational environment
Owner	EURECAT		

Table 4 AI Resource factsheet "Dynamic Planner & Scheduler"

4 How to Access the Toolkit

As part of the AI REGIO Network Offer, the AI4Manufacturing Toolkit will be accessible to our final user via the AI REGIO DIH Service Marketplace that will be a centralize hub for the providers where they can house knowledge, expertise and technology to support companies with piloting, testing and experimenting with digital innovations and, for consumers (e.g., SMEs) a one-stop-shop able to offer multiple services.

The current version of the AI REGIO DIH Service Marketplace[2] is already publicly available at https://airegio-portal.eu/, even if the user registration is not open.

At a later stage and when the content population will consolidate the initial collections of assets and knowledge and the content is enough to highlight to our final user expectations for the AI REGIO Marketplace future success, we'll open the registration also to our final user.

The first version of the workspace structure has been designed to meet the identified requirements coming from the AI REGIO stakeholders and includes different sections: One of those is entirely dedicated to the AI4Manufacturing Toolkit. Starting from the main AI REGIO Marketplace landing page, our Platform user will be able to access the toolkit from the dedicated link or menu button that will direct them to the **AI4Manufacturing Toolkit Dashboard** useful for displaying the toolkit items organized into the three main identified categories (AI resources, AI pipelines or AI orchestrator).

[2] For more information refer to D3.7 AI REGIO DIH Service Marketplace and D4.7 AI REGIO DIH Platform

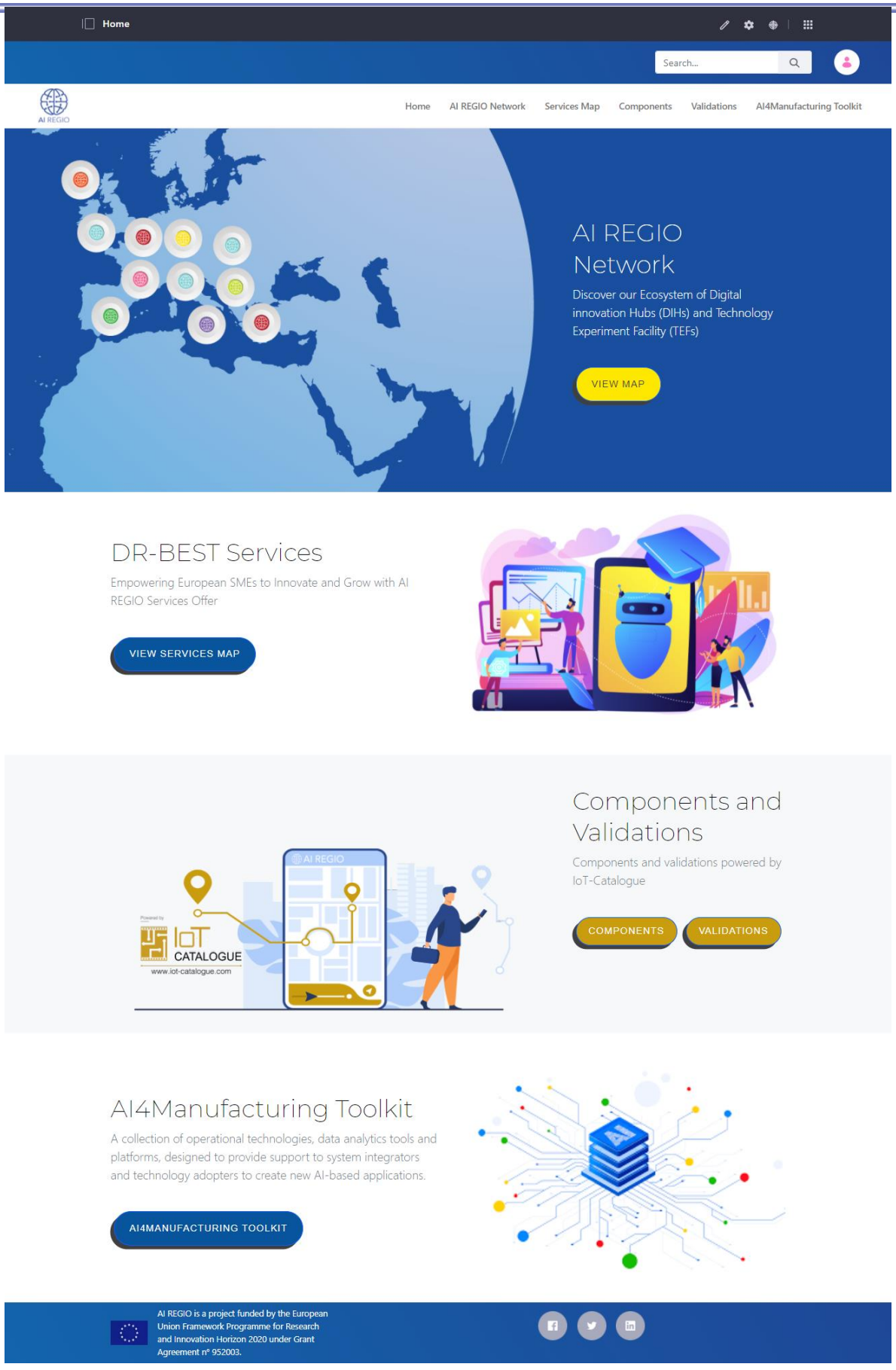

Figure 3 AI REGIO Marketplace Landing Page

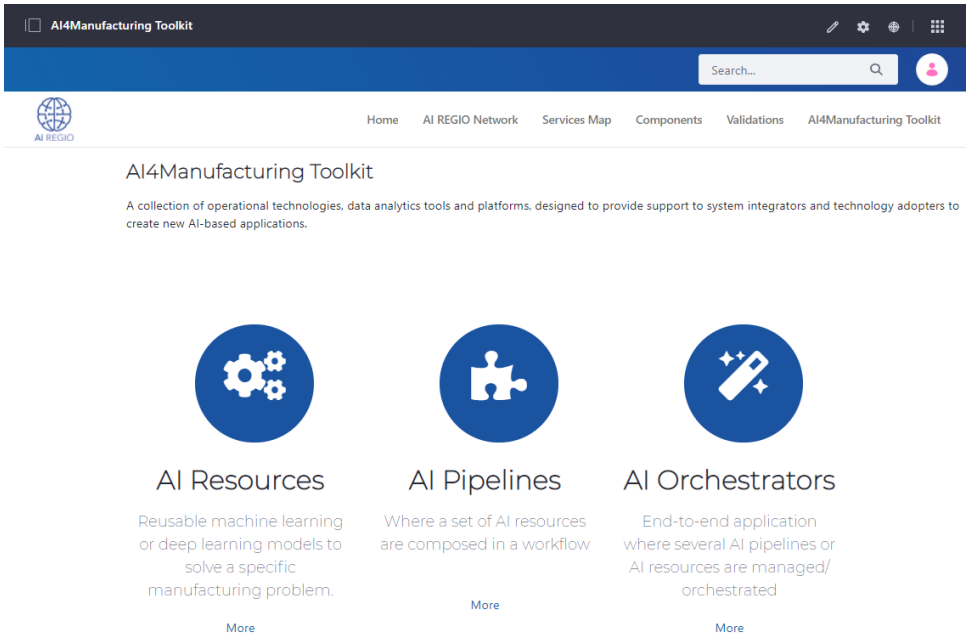

Figure 4 AI4Manufacturing Toolkit Dashboard

5 Conclusions and Next Steps to D4.6

This deliverable focuses on AI4Manufacturing Toolkit and it is related to task WP4.3. As explain in previous sections, currently AI4Manufacturing Toolkit is a collection of operational technologies, data analytics tools and platforms, classified between AI resources, AI pipelines or AI orchestrator. For each asset a factsheet has been generated, presenting foreground to be developed within AI REGIO project and, as examples, three specific factsheets (one from an AI resource, one from an AI pipeline, and one from an AI orchestrator) have been included in this document. Besides that, the AI technological assets catalogue is currently accessible through a specific landing page in AI REGIO portal.

In the coming months this catalogue of assets will be extended with new AI assets coming from the experiment and open calls, among others. Moreover, second release of the Toolkit will be oriented to semantically enrich the categorization, offering the end user not only a catalogue but also making it easy to adopt its content. In that sense some initial ideas have been put on the table, inspired by both the categories used by the AI4EU platform and other categories closer to the manufacturing domain.

Furthermore, as added value, it has been proposed evolved the current Toolkit towards a Platform where end users are able to try/play with the AI4Manufacturing's tools. Since the designer environment where each tool was created is different, a feasibility analysis will be carried out in order to define how this execution environment could be launched.

Main conclusion of all these proposed activities will be summarized in the deliverable D4.6 that accompanying the second iteration of task WP4.3. That document will be submitted on M30.