self-X Artificial Intelligence for European Process Industry digital transformation

Deliverable

D2.1 Scenarios and Requirements for Self-X AI adoption in Process Industry

Deliverable Lead: CARTIF

Deliverable due date: 31/01/2023 (M9)

Actual submission date: 14/07/2023

Version: V2.0

Funded by
the European Union

Document Control Page	
Title	**D2.1 Scenarios and Requirements for Self-X AI adoption in Process Industry**
Lead Beneficiary	**CARTIF**
Description	**Analysis of the industrial Scenarios and the preliminary design of the requirements from use cases of the project for adoption in Process Industry of self-X or autonomic AI**
Contributors	**AIMEN, BFI, CARTIF, CORE, DEUSER, EIFFAGE, ENG, IDALSA, MSI, NISSA, POLIMI, RCPE, SID**
Creation date	**21/06/2022**
Type	**Report**
Language	**English**
Audience	☒ **public** ☐ **sensitive**
Review status	☐ **Draft** ☒ **WP leader accepted** ☒ **Coordinator accepted**
Action requested	☐ **to be revised by Partners** ☐ **for approval by the WP leader** ☐ **for approval by the Project Coordinator** ☐ **for acknowledgement by Partners**

Version	Author(s)	Changes	Date
0.0	CARTIF	TOC	21/06/2022
0.1	CARTIF, EIFFAGE, DEUSER	Update on Asphalt Use Case	25/11/2022
0.2	SIDENOR, MSI, BFI	Update on Steel Use Case	02/12/2022
0.3	RCPE	Update on Pharma Use Case	02/12/2022
0.4	IDALSA, AIMEN	Update on Aluminium Use Case	13/12/2022
0.5	EIFFAGE, NISSA, CARTIF	Update on Asphalt Use Case, Section 5, Conclusions section	16/01/2023
0.6	AIMEN, CARTIF	Update on Aluminium Use Case, Sections Integration	19/01/2023
0.7	SIDENOR	Revision of the unified doc 0.6	24/01/2023
0.8	CARTIF	Revision after 1st reviewer	25/01/2023

0.9	NISSA	Update Section 5	26/01/2023
1.0	CARTIF	Final Review	30/01/2023
1.1	CARTIF	New template and indications	15/05/2023
1.2	EIFFAGE, DEUSER, CARTIF	Update on Asphalt Use Case	30/05/2023
1.3	AIMEN	Update on Aluminium Use Case	30/05/2023
1.4	RCPE	Update on Pharma Use Case	31/05/2023
1.5	SIDENOR, BFI	Update on Steel Use Case	06/06/2023
1.6	CARTIF	Integrated version with all contributions	20/06/2023
1.7	CARTIF	Final version for review	30/06/2023
1.8	Alejandro Cuadrado de la Oza (External Ethics Advisor)	Ethics issues compliance review	03/07/2023
1.9	SIDENOR	Internal review	10/07/2023
2.0	CARTIF	Final version	14/07/2023

Disclaimer

Table of Contents

List of Figures

List of Tables

List of Acronyms

Aluminium	AL
Artificial Intelligence	AI
Asphalt	AS
Creative commons	CC
Data Management Plan	DMP
Deliverable	D
Description of the action	DoA
Electric Arc Furnace	EAF
European Commission	EC
Functional	F
Industrial Internet of Things	IIoT
Key Performance Indicator	KPI
Knowledge Discovery from Data	KDD
Message Queuing Telemetry Transport	MQTT
Non-Functional	NF
Pharma	PH
Project Month	M+"number"
Quality Control	QC
Reclaimed Asphalt Pavement	RAP
Reinforcement Learning	RL
self-X	s-X
self-X Requirements	sXR
Steel	ST
Structured Query Language	SQL
System Requirements	SR
Task	T
User Interface	UI
User Story	US
Work package	WP

Abstract

The s-X-AIPI project tries to optimize industrial processes through artificial intelligence technology, based on the analysis of initial data and machine learning technologies, which appear thanks to the reception, analysis and translation of new data. Once new data are received, thanks to the continuous monitoring of the process, these are added to the initial database, increasing its experience and learning new formulas, being progressively capable of offering more optimal parameters in order to achieve a better and efficient process of higher quality and lower cost.

The fact of increasing performance and efficiency of the different industrial processes, may be affected depending on the industrial context. Due to this reason, this document examines the different scenarios found in each use case worked on this project: asphalt, aluminium, steel and pharmaceuticals. So current deliverable D2.1 proceeds to describe the identified requirements to achieve applications of artificial intelligence focused on the optimization on processes of the different use cases.

The requirements contain aspects of the product, the process, the quality of the data, the traceability of the product carried out throughout the entire process, the diagnosis of the plant, etc. Everything contemplated for improvement of process, helping the human in the loop to make decisions to obtain better results. In addition, technical requirements to achieve these objectives are described below. It also describes the different necessities related to the infrastructure of the facilities, which must be equipped with good monitoring equipment for data capture, collection and storage.

Scope

Deliverable D2.1, "***Scenarios and Requirements for Self-X AI adoption in Process Industry***" is the result of task T2.1 "***Requirements for self-X AI solutions collaborating with humans***" within WP2, "***Design and Architecture of self-X AI solutions integration in process industry plants***" of the project.

Work Package 2, led by partner POLIMI, involves four deliverables:

- D2.1 Scenarios and Requirements for self-X AI adoption in Process Industry. (Leader: CAR)
- D2.2 AI for Process Industry Reference Architecture and implementation (Leader: ENG)
- D2.3 AI Maturity Model for Process Industry (Leader: POLIMI)
- D2.4 Guidelines for trustworthy AI in process industry (Leader: CAR)

The partners that have contributed in the definition of the requirements of the self-X solutions are those who are working on the different use cases of the project, each one contributing according to their role. The description of the participants in this document can be found below. Nevertheless, this deliverable is addressed to the whole s-X-AIPI consortium, due to the requirements associated to self-X and KPIs that are important for the development of the next Work Packages.

The different roles that partners contribute to current D2.1 are the following:

CAR, EIFFAGE and DEUSER contribute on the asphalt use case scenarios and requirements.

SID, MSI and BFI work on steel use case defining their specifications and requirements.

RCPE is in charge of the pharma use case.

IDALSA and AIMEN are in charge of aluminium case.

In addition, and alongside the rest of the partners, ENG, NISSATECH, CORE and POLIMI act as transversal partners for all use cases, contributing and assisting in the definition of the different requirements for the self-X tools requirements.

The scope of this deliverable D2.1 is within Task 2.1 where the initial description of requirements have gone through months M2 to M9. Additionally, a second period of analysis of requirements has been contemplated in this project, with the aim of redefining, if necessary, the requirements of the elaborated self-X architecture. In the event of any modification of the requirements described in the current document, these ones will be reflected in deliverables of the following WP.

The guidelines of the s-X-AIPI project are only applicable for this project and refer specifically to use cases from this project.

As a summary, this deliverable D2.1 is devoted to generate the corresponding requirements (functional and non-functional) of self-X abilities for AI procedures and components for the AI-based applications. The outcome of this deliverable will also help define how autonomic managers will coordinate the different elements of the AI Data pipeline and interaction (via UI) with the different human roles in each use case scenario. Requirements also will define: KPIs (socio-environmental-business-technical) affected to later evaluate them in demonstration phase in addition of the availability of data sources and particular IT systems to interact with the "Data ingestion" component of the self-X AI data pipeline.

1 Introduction. Deliverable Description

1.1 Context of deliverable

This document describes a preliminary design of the requirements for adoption in Process Industry of self-X or autonomic Artificial Intelligence based on different use cases of the project.

The detailed requirements have been drawn up after carrying out a deep analysis of the European industrial context and in particular of the scenarios found in the use cases that concern s-X-AIPI project, steel, aluminium, asphalt and pharmaceutical.

The objective of this Deliverable is to generate functional and non-functional requirements of self-X abilities for AI procedures and components for AI-based application in order to collaborate with humans, coordinating elements of AI Data pipeline and interacting (via user interfaces) with human roles, as well as defining technical aspects, environmental and business Key Performance Indicators affected, which are analysed in order to later evaluate them in the demonstration phase. Another requirements take into account the availability of data sources and the available IT systems to interact with "Data ingestion" component of the named self-X data pipeline.

1.2 Relationship with other tasks and deliverables

This deliverable together with those that accompany this Work Package (WP2) are the basis for the following Work Packages.

The definition of the requirements establishes the objectives of the self-X architecture to be developed.

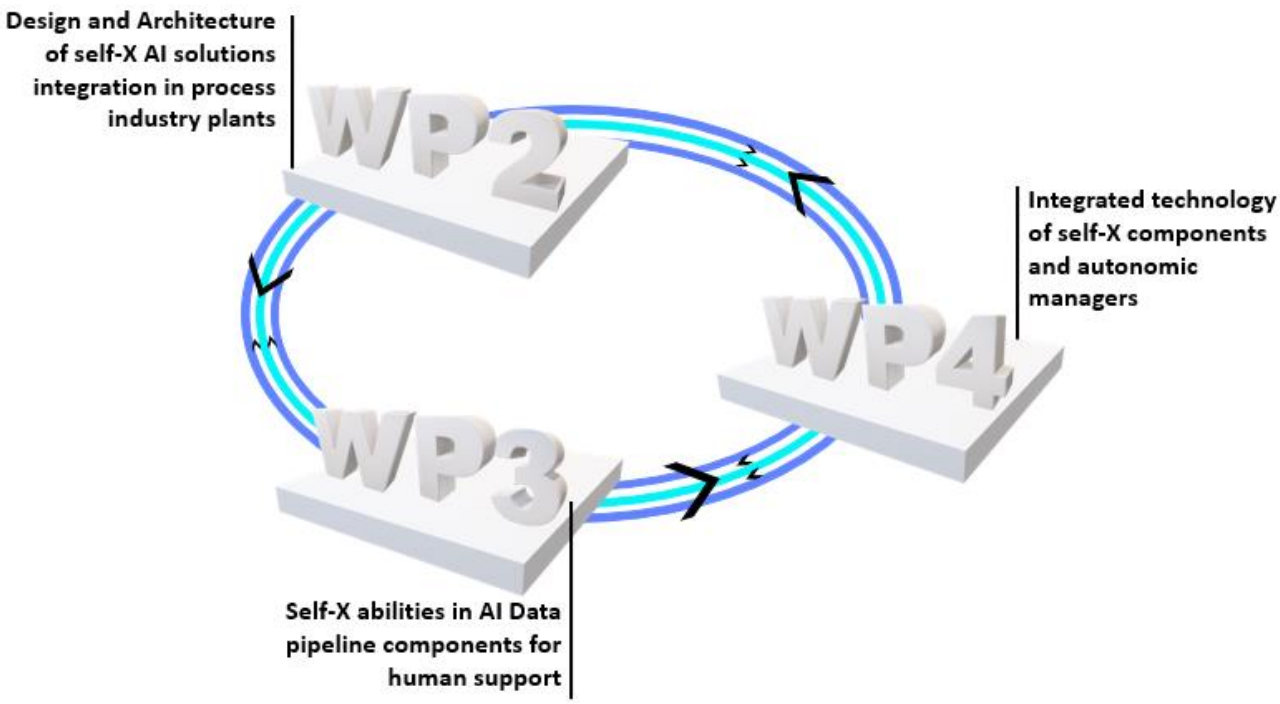


Figure 1. Relationship between work packages

2 User Scenarios Description and Expectations for process industry

2.1 Asphalt

2.1.1 Objectives/Benefits

EIFFAGE asphalt use case focuses on circularity of the value chain (from quarry to road) with respect of the quality control of feedstock (aggregates, bitumen, recycled asphalt, ...) the overall sustainable performance of the process (including asphalt paving) and the quality of final product (asphalt mix).

Application of toolset to Asphalt: AI application for integrating value chain information from quality laboratories, production and paving logistics and improve overall sustainable performance of the process by including human feedback for asphalt mix design adaptation, maintenance actions and quality monitoring.

The objective is to create self-X AI applications for whole value chain to gather and infer information from all steps of Asphalt manufacturing with data from: (1) Production, (2) Laboratory, (3) Paving and logistics (asphalt transport to the job site) and (4) Raw materials (aggregates, bitumen, Reclaimed Asphalt Pavement or RAP and additives). To improve the efficiency of the whole value chain, AI apps will be used in predictive diagnosis of equipment efficiency to infer energy savings in production plant and maintenance optimization.

2.1.2 General Description

Asphalt production needs bitumen, aggregates, and additives. As time passes, old asphalt pavements have to be removed, and asphalt plants have started to reuse the reclaimed asphalt pavement, or RAP. To complete the circularity economy process, it is needed a high-quality control of the process along the value chain (from quarries and petrochemical companies to paving asphalt at the road or job site including the logistics). Major goal for asphalt companies is to increase this RAP volume (RAP rejuvenators based on ecological oils are also used to soften this recycled material and help increasing the quality mix) and use more additives from other recycled materials (e.g., fibres and crumb rudder from End of Live Tire).

Figure 2. Asphalt use case value chain

This increased use of recycled materials in asphalt mixes needs adaptation of processing technologies and obtained final mix will depend on those used materials and process parameters among different plants: type of baghouse, used fuel, dryer control, plant operator... Changing one of these parameters results in different mixes quality, fuel consumptions and CO2 emissions. EIFFAGE Plants owns different platforms to track production data (incl. paving logistics operation) and laboratory quality data. One main objective is to connect those data sources so final asphalt mix can be analysed, from cold aggregates to final paving, and link quality problems in one process stage and the rest with the help of the AI. Production data from the plant alongside asphalt data from transport (logistics IIoT solutions from original quarry to asphalt paving and extension at job sites), including final asphalt mixes and RAP data from laboratory data will be used to cover whole value chain of asphalt production. Current quality process checking is done in four stages, (1) raw materials analysis, (2) analysis of use of materials at plant, (3) analysis of mix production and (4) analysis of different parameters related to onsite activities or works (asphalt mix laying and compaction). The delay of the process can be 30 days from raw materials analyses to final mix analysis.

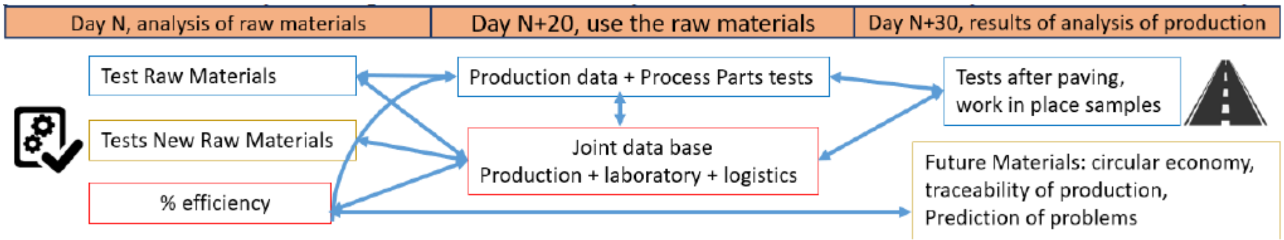


Figure 3. Laboratory analysis flow

2.1.3 Current situation (i.e. operating modes)

2.1.3.1 Asphalt plant

By weight, 92-96% of asphalt consists of course aggregates, sand, and filler – aggregates that give asphalt its strength. The remaining 4-8% comprises binder that binds all these materials together with some additives. That binder is usually bitumen derived from crude oil. A scheme of an asphalt mix batching plant is shown in the following figure:

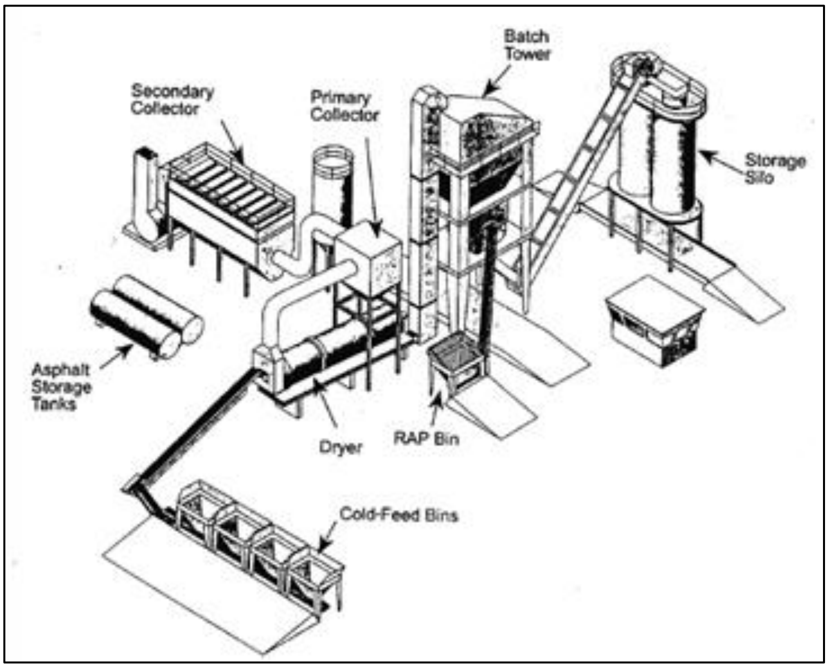


Figure 4. Asphalt mix batching plant

Asphalt mix production is a sequence of processes where aggregates are blended, heated, dried, screened and mixed with binder in designated amounts to produce an asphalt mix, meeting specified requirements concerning performance parameters according with its intended use. The process begins when the stockpiled aggregates in the cold feed are metered and conveyed to a dryer drum where they are heated to a specific temperature. A first collector removes large dust particles from the gases before entering the bag house, which removes fine particulate matter before they are released into the atmosphere. At the final stage of the process, the mixing unit (mixer) receives aggregates, bitumen, filler, even RAP and additives from their weigh hoppers, and then, will mix all the contents thoroughly for a fixed time before discharging the contents into the truck or into storage silos.

In EIFFAGE Atalaya (Seville) asphalt plant, which is the asphalt plant where the asphalt use case demonstrator is located, a pro-software solution that is fed by data from automata and sensors was installed. This solution, which is called "Connected Plant", collects production data to facilitate inventory management and supervises the quality of asphalt mixtures, the consumption of energy and environmental impacts.

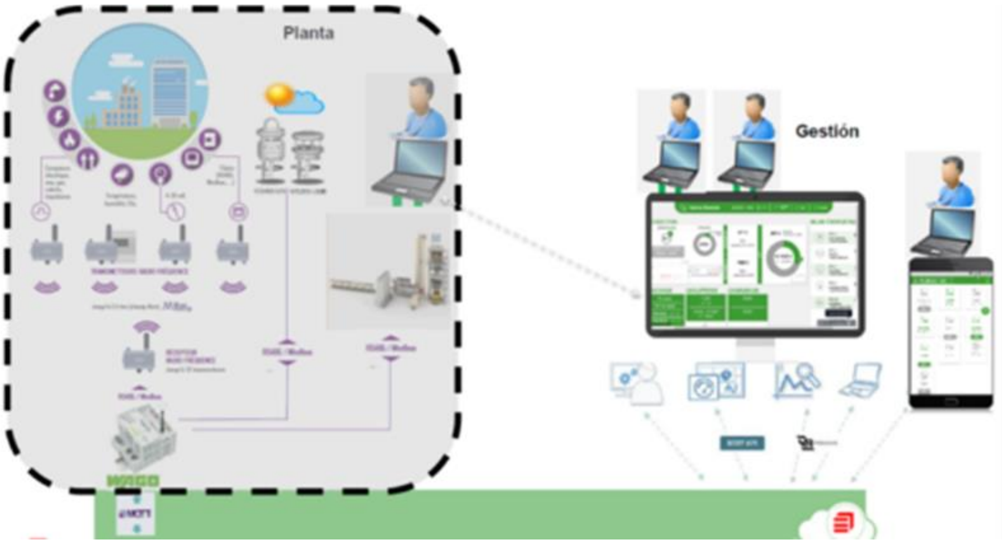


Figure 5: EIFFAGE Connected Plant solution.

2.1.3.2 Asphalt mix transport and application

Once hot asphalt mix is produced, further transport is required to the job site where it is laid down and then compacted. Asphalt temperature while laying and compacting, significantly influences the durability of asphalt pavement. So, collected information should be used to ensure target properties, also subject to national or international specifications. Compaction process significantly influences asphalt layer surface properties, which is an important factor in noise reduction, water splash and skid resistance.

Figure 6: Asphalt production, transport and application.

In this regard and, related to asphalt mix transport and onsite activities, EIFFAGE has developed some technologies to provide information along asphalt laying and compaction process to improve it and enhance human operator decision-making. The information is collected to be used, to ensure that the target properties of the layer are more uniform and to improve the efficiency of the process. A series of web services have been integrated that collect the information sent by the sensors and store it in a SQL Server database. This solution provides efficient and reliable data management that offers a complete data warehouse.

2.1.3.3 Laboratory: Quality Control (QC)

The role of the Quality Control Laboratory for the Asphalt Plant is to ensure that the asphalt mixes produced by the plant meet the required specifications by the current specifications or customer requirements. The Control Laboratory Technicians communicate with and assists plant personnel to ensure the best possible quality product. On one hand, the mix design consists of proportioning the aggregates and selecting the optimum bitumen content so that satisfactory air voids are obtained. In some cases, performance tests are conducted in the lab to help ensure desired mix performance is obtained.

It is essential to ensure that the plant produced mix meets the properties of the designed mix. Even though satisfactory properties are obtained with materials mixed in the lab, it may take some adjustments in the mix components to satisfactorily meet the mix design requirements during production. If the mix properties are not satisfactory during mix production, then adjustments of the mix components are required. On the other hand, products production quality control is performed taking mix samples from loaded trucks, but this must be done very carefully, taking a representative sample from the truck. The sampling frequency must be done according to CE marking requirements.

Figure 7: EIFFAGE Laboratory Quality Control

In Atalaya (Seville) asphalt plant, EIFFAGE own laboratory works with a comprehensive management system called “hclab”. “hclab” is a specific management system for laboratories involved in construction quality control and material testing. It encompasses the technical processes of the asphalt plant (sample registration, testing, report generation, quality, etc.). This system allows a reliable and traceable data flow related to information of the lab. Also, “hclab” allows the design of custom tests or modify -if needed- any of the tests available in its library.

Figure 8. EIFFAGE laboratory system

Considering the previous stated information, we could summarize the current situation as follows:

Table 1. Operating current situation modes

Platform or system	Data availability	Involved users	Objectives	Performance indicators
Asphalt mix production plant	Asphalt mix production data (composition, process parameters, etc.)	Asphalt plant manufacturing operators, foremen and engineers.	Info about asphalt mix composition and manufacturing process parameters.	Deviations in composition from the asphalt mix designed and, deviation from process optimum parameters or equipment performance.
Transport and paving	Asphalt mix temperature, truck's location, and shipment, etc.	Asphalt paving operators, foremen and engineers.	Info about asphalt mix and weather conditions at job site (Temperature, shipment time and application time, etc.)	Deviation from optimum temperature for paving or compaction and, weather condition limitations.
Laboratory	Asphalt mix composition, volumetric and mechanical properties data.	Operators and laboratory technicians, QC engineers.	Asphalt mix quality test parameters according to standards.	Establish meaningful, but achievable, specification values (e.g., performance test thresholds) that consider the overall test variability.

2.1.4 Future situation: improvements with self-X solutions

The general objective is to create self-X AI applications for whole value chain to gather and infer information from all steps of Asphalt manufacturing and application with data from:

(1) Production,

(2) Laboratory,

(3) Paving and logistics (asphalt transport to the job site) and

(4) RAP and additives.

For the asphalt use case, it is expected the creation of AI procedures with self-X capabilities (detect, diagnose and repair) for integrating the data from quality laboratory, production data and paving logistics. It involves AI procedures for data ingestion, transformation and exploration followed by an initial set of AI models (mainly clustering and ML). AI solutions initial research and development for the asphalt use case will be based on available historical data.

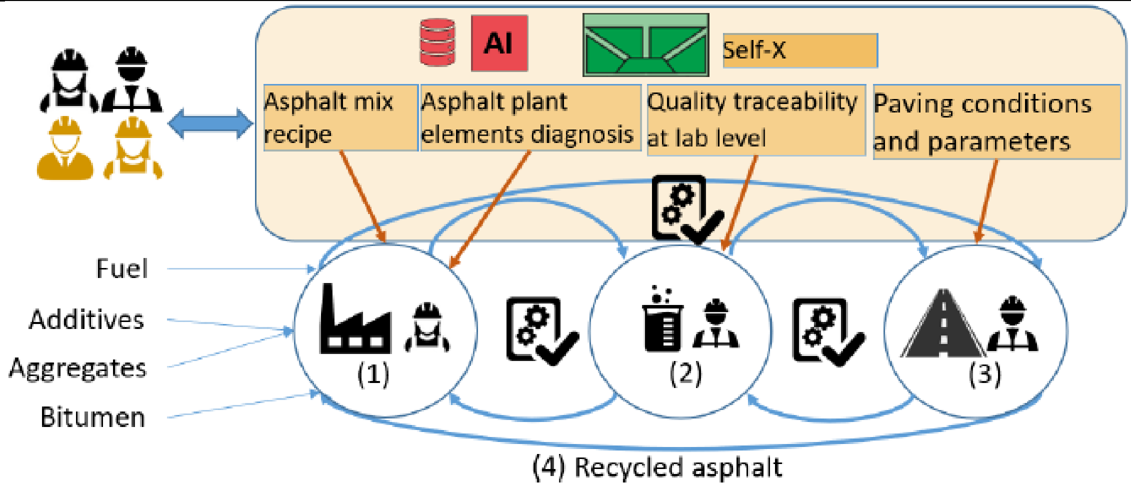


Figure 9. Self-X Asphalt use case

Within asphalt use case next technical objectives will be addressed, with the corresponding AI solutions:

Table 2. Self-X capabilities

AM[1] capabilities	SELF-X detect	SELF-X diagnose	SELF-X repair	Feedback for human in the loop	Contributions
Asphalt mix design	of asphalt mix design aggregates and additives' quantities	Asphalt mix according to indicated design		Need of asphalt mix design adaptation	Process, Circularity
Plant elements diagnosis	of asphalt plant performance elements KPI's	source of future failures & need of maintenance		Reports possible maintenance to be performed	Process
Paving conditions and parameters	of adequacy of different mixes paved at job site	possible modifications in the asphalt mix design	Change / adaptation of product mix design	Modifications of current asphalt mix design to adapt to paving site conditions	Process
Quality traceability at lab level	Quality control of final mixes al lab level	Asphalt mix quality adequacy according to standards		Indication of quality parameters according to desired KPI's	Sustainability

2.1.4.1 AM capabilities: asphalt mix design.

Main objective within this self-X solution is focused on the asphalt mix design at plant level. To reach it, the following specific self-X AI solutions have been proposed for development.

1. Knowledge Discovery from Data, taking into account all the production data currently registered.

[1] Autonomous Manager.

2. Smart visualization dashboards to monitor both, the most important variables and key process parameters of interest.

Figure 10 represents the value chain for asphalt mix design where domain expert belongs to Asphalt plant.

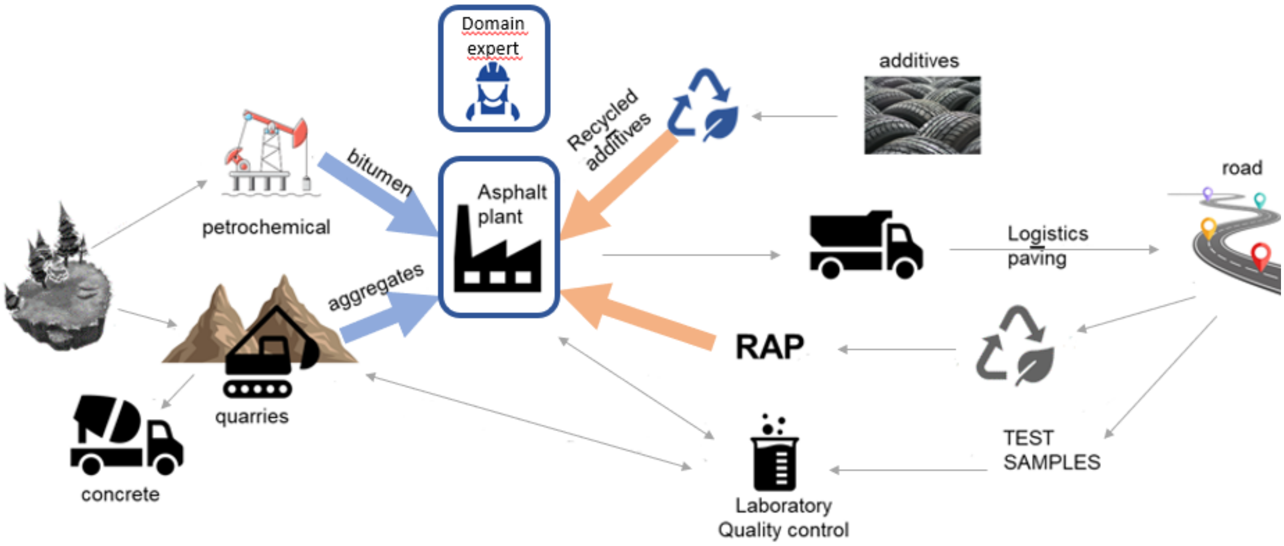


Figure 10. Value chain for asphalt mix design

From production data, asphalt mix composition deviation and key process parameters deviation from every batch production happens due to inevitable systematic and random errors. To reduce the deviations of the component content and production parameters as much as possible, all available production data will be taken into account and several multivariate analysis, outlier's detection and/or other supervised and unsupervised intelligent algorithms to predict unusual/unlikely situations will be carried out.

2.1.4.2 AM capabilities: asphalt mix plant and elements diagnosis.

Main objective within this self-X solution is focused on plant elements diagnosis. Predictive maintenance uses historical and real-time data from various parts of the operation to anticipate problems before they happen. The real-time monitoring of asset condition and performance factor into predictive maintenance. Figure 11 represents the value chain in the Asphalt plant.

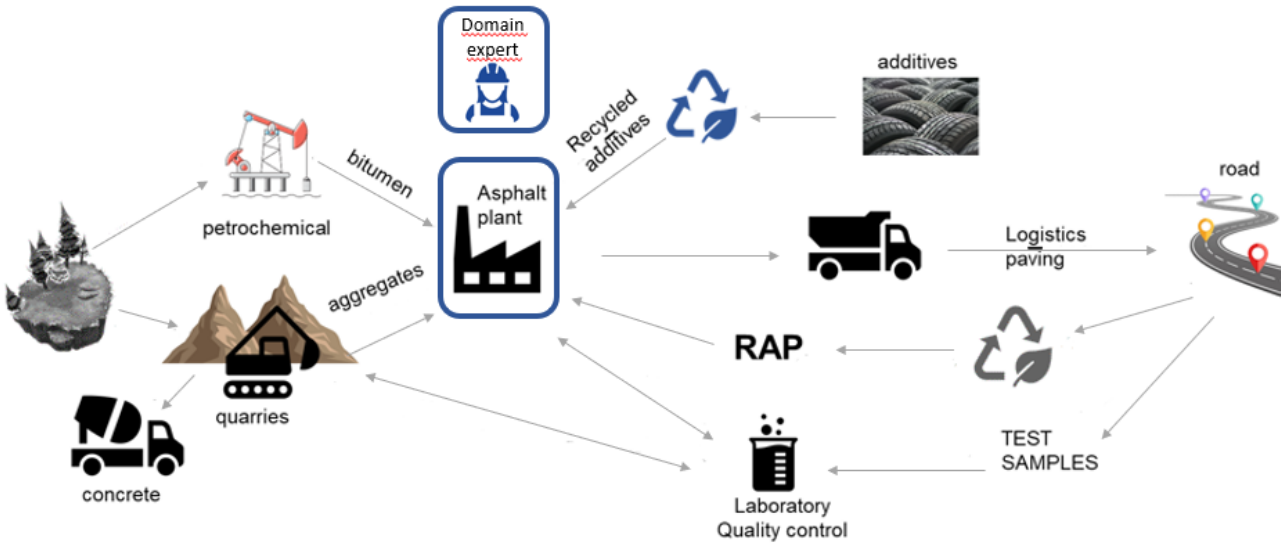


Figure 11. Value chain for asphalt plant

AI solution for plant elements diagnosis will focus on self-healing method capable of responding to self-diagnose and unexpected situations, ensuring predictive maintenance (PM) of engines of the production chain.

Here, the aim of the s-X-AIPI project is to create a framework capable of providing advanced solutions (including AI solution provided by DEUSER) to improve situational awareness by visualizing real-time raw data and prediction by analysing those data, as well as to detect real-time anomaly, change or drift (self-healing related) to support human or AI-powered (self-AI capacity) decisions, and improve organizational and operational capabilities for maintenance tasks.

By leveraging the WP3 results, the Asphalt use case and in particular the diagnosis solution will first contribute to piloting self-healing solutions in a realistic environment and the self-recovery mechanisms that lead to effective prediction will be explored. Recovery (stabilization) actions can be done via the activation or modification of appropriate parameter such as adjust variator (variable frequency drive), activate ventilator, or cleaning action.

Technology will be provided for automating production continuity and recovery actions allowing to discover in advance anomalies or alarms and automatically activate, in the foreseen cases, the relative maintenance actions before situations can worsen. This will allow anomalies detection and decrease the response time (KPI EIFFAGE) and will be able to prevent the deterioration of service through the prior activation of the appropriate maintenance response.

2.1.4.3 AM capabilities: paving conditions and parameters.

Main objective within this self-X solution is focus on the paving conditions and parameters. To reach it, the following specific self-X AI solution has been proposed for development:

1. Laying and compacting temperature prediction model, based on conditions of the asphalt mixture when it leave the production plant and when it is spread out on the road.

In Figure 12, the value chain for '*logistics paving*', where domain expert belongs to Asphalt plant and Job Site, is presented.

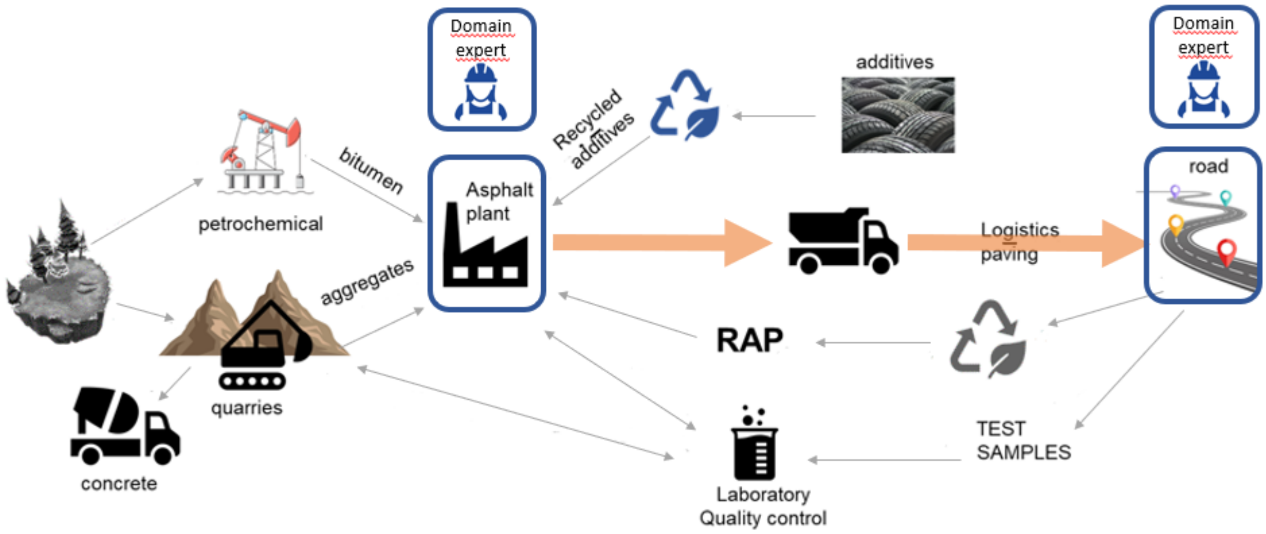


Figure 12. Value chain for logistic paving

In that sense, considering as main data set both the paving and logistic conditions monitoring system and the production plant control system, several data analytics techniques will be applied. The aim is to explore the data and generate models for prediction, in order to develop dashboard able to propose improvement within the asphalt mix design.

2.1.4.4 AM capabilities: Quality traceability at lab level.

Asphalt mixture is composite material consisting of different content of asphalt binder, aggregates, filler, air voids, and, in some cases, additives. Its constitutive mechanical performance is significantly influenced by these constituents at different scales and their contents in the mixture. When asphalt mixes are produced in asphalt mix plant, the component content always deviates from the asphalt mix design or from the designed formula due to inevitable systematic and random errors. The larger the deviations of the component content in the produced asphalt mix, the worse quality obtained.

Main objective within this self-X solution is focus on the quality traceability at laboratory level. To reach it, the following specific self-X AI solutions have been proposed for development:

1. Verify that the results obtained in the laboratory tests (done on a frequency predefined according production rules) conform to what is established in the asphalt mix design (coming from production data) and comply with the standards.
2. Develop a volumetric/mechanical properties prediction model based on gradation properties of the mixes (coming from production data and laboratory test).

The value chain for 'Asphalt Laboratory Quality control' is represented in Figure 13 where the domain expert belongs to Laboratory.

In that sense, considering as main data set this obtained from the laboratory, and fusing it with information available from the asphalt mix design (production data), several data analytics techniques will be applied to explore the data and generate models, in order to develop dashboard able to support the laboratory operators in their daily activities.

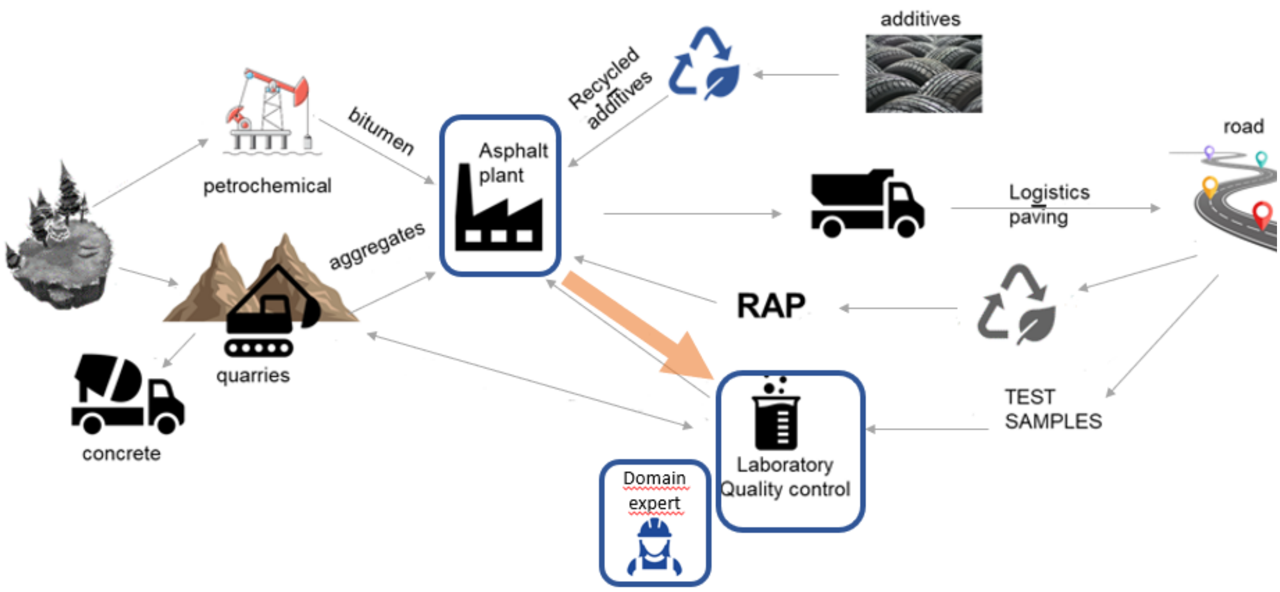


Figure 13. Value chain for asphalt laboratory quality control

2.1.4.5 Human roles interacting with self-X solutions.

In order to improve the self-X proposed in this project, it is necessary to include some feedback from human that provides additional information to the given predictions. Different interventions are contemplated, ranging from the introduction of previous data for the correct functioning of the solutions, to the final validation or a return of data obtained or modified thanks to the self-X.

The contribution of Human in the loop remove it is very important, these solutions are improvements to the process that facilitate the work of the operators, but without forgetting that their contributions will improve the predictions obtained. From these interactions it is expected:

- Real-life feedback to the final AI applications

- Labelling for update model training
- Drive the artificial data generation
- User interface interaction with Autonomic manager
- Actions of the human, based on support of the AI application.

Modifying the self-X pipeline as follows:

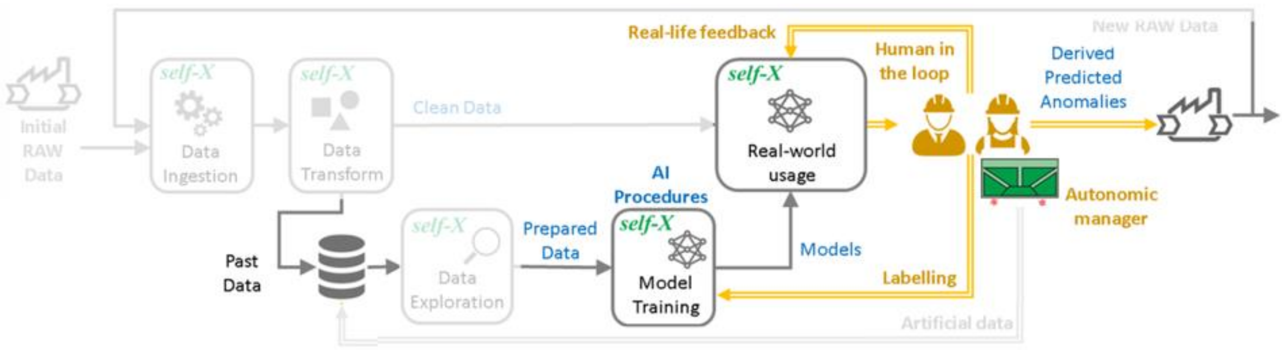


Figure 14. AI data pipeline proposed with integration into manufacturing, with human in the loop and self-X autonomic managers for reduced human intervention (connected with double line arrow)

The interaction has a moment and a purpose for each solution.

Table 3. Human in the loop Asphalt Use Case

Self-X Asphalt solution	Impact to human in the loop	Feedback from human in the loop
Asphalt mix design	Improve the available knowledge about the asphalt production process from data monitoring	Label and/or Validate the knowledge discovered in the data, i.e. assigning a percentage of confidence that will be used later to improve the autonomic manager
Plant elements diagnosis	Review of the elements of the plant according to the alerts obtained, planning of predictive maintenance proposed by the s-X	Document the maintenance interventions carried out, breakdowns and reasons for them. Document and validate predictions
Paving conditions and parameters	Control and modification of asphalt mix conditions that leaves the plant in order to ensure the quality required in the job site	Verification of the quality and temperature from asphalt mix in job site increasing data ingestion
Quality traceability at lab level	Control and modification of asphalt mix design in order to obtain the optimum mix according to its requirements	Validation of the predictions obtained

2.1.5 KPIs: Baseline values and expected Results

KPI Category	Code from DoA	KPI definition	Criticality	Proposed formula/unit	*Baseline*	*Target value (%)*
Business KPIs	KPI1.1_Improvement of Productivity	Productivity increased in terms of energy efficiency kWh/ton and production costs €/ton by optimizing production lots. – (1) Improvement of logistics factory-jobsite. – (2) Saving raw materials (increasing recycling). – (3) Decreasing rejections of intermediate materials, i.e., aggregates heated and not used by optimization of recipes and cycles of batch manufacturing.	High	Comparison among asphalt mix compositions: 0% RAP, 10% RAP, 20% RAP, 30% RAP, 40% RAP, etc. Formula = kWh/Tn (kWh = Electricity consumption (kWh) + Fuel (Kg * 11.2 kWh/kg) + (Gasoil = L * 10.96 kWh/L) / (Tn Asphalt mix) [kWh/Tn per asphalt type]	92.4	87.8 (-5%)
Technical KPIs	KPI1.2_Imrpovement of quality	Improved asphalt mix (reducing deviations in quality) – (1) better adjustment of filler/bitumen rate. – (2) reducing deviations in dosing process of the plant. – (3) better monitoring/control of effective hot aggregates separation (screening).	Medium	(1) Filler/bitumen rate (per asphalt mix type) = coefficient (vs. specifications); (2) Dosing deviations (per asphalt mix type) = real quantity of aggregates, bitumen and filler / specified quantity of aggregates, bitumen and filler (%); (3) Models (per asphalt mix type): gradation, volumetric parameters, and mechanical parameters deviations.	N/A (According to product specifications, CE marking, etc)	
Technical KPIs	KPI1.3_Improvement of response-time	Reduction of intermittent downtimes	Medium	kWh/Tn according to KPI 1.1 / (Number of daily stop/start) [(kWh/Tn)/NSS]	18.5	17.6 (-5%)

Environmental KPIs	KPI2.1_Reduction in generated CO2 emissions	Savings of electric energy, heavy fuel, and diesel account for 15.5kWh/t of asphalt produced it is equivalent to a reduction of CO2 in 5.44 kg/t. Calculation of energy spending and CO2 reduction for asphalt mix ton according to asphalt mix composition, weather conditions, raw materials characteristics, etc. CO2 emissions per ton of asphalt mix manufactured.	High	Electric energy (100% renewable = 0.0 gr CO2 / kWh); Heavy fuel: CO2 = Litre of Fuel (L) * 3.110 gr CO2; Gasoil (diesel): CO2 = Litre of Gasoil (L) * 2.705 gr CO2 [kgCO2/Tn asphalt]	23.5	18.1 (-23%)
Environmental KPIs	KPI4.3_Reduction in use of resources: bitumen	Circular manufacturing and reducing CO2 emissions and waste in asphalt manufacturing. Tons of bitumen reused.	High	Total aggregates Tns - (Tns of RAP * 0.96) [Tns/Y] [2]	96000	72000 (-25%)
Environmental KPIs	KPI4.3_Reduction in use of resources: virgin aggregates	Circular manufacturing and reducing CO2 emissions and waste in asphalt manufacturing. Tons of aggregates reused.	High	Total bitumen Tns - (Tns of RAP * 0.04) [Tns/Y] [2]	4000	3840 (-4%)
Social KPIs	KPI5.3_Increase of competence, skills and qualification	More attractive and safer Jobs	Low	Survey to operators	N/A	4/5 acceptance

[2] Estimated Asphalt mix production of 100.000 Tn

2.2 Steel

2.2.1 Objectives/Benefits

Steel use case focuses on the optimised use of scrap (raw material to manufacture steel) to produce high-quality steel products, avoid downstream surface quality problems that leads to scrap material and reduce required energy intensity (mainly during scrap melting in the Electric Arc Furnace EAF stage). The improved control of varying steel scrap properties allows:

- Tighter supervision of scrap suppliers
- Support for enhanced usage of low-cost scrap types
- Improved process end-point control
- Homogeneous production by shared knowledge

Therefore, the s-X-AIPI project is proposed for improving:

- The control in the plant by using Machine Learning for calculating the scrap chemical composition, which is normally affected by disturbing elements like copper, sulphur or phosphorus, considered as residual elements which downgrades the steel quality, and with high degree of uncertainty.
- The scrap mix that minimizes the cost and meets the customer requirements in terms of liquid steel chemistry.
- Assure the best performance of the EAF process for achieving the required steel temperature and chemical composition.

2.2.2 General Description

The s-X-AIPI solution for the steel case will be validated in Sidenor's Basauri steel plant as a pilot demonstrator. Sidenor's plant has facilities for handling steel scrap (scrap yard), melting at Electric Arc Furnace (EAF), secondary metallurgical treatments (refining), solidification at continuous casting, hot rolling and finishing of the steel products. More precisely, the project and solutions are mainly focused **from the scrap yard to the EAF process** prior to the refining stage.

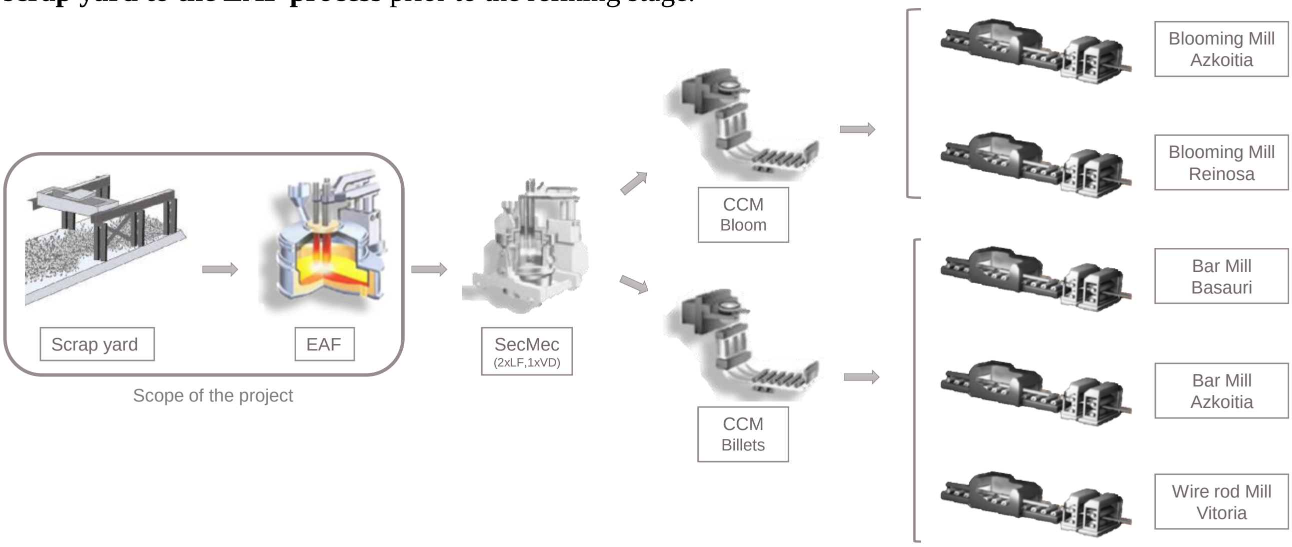


Figure 15. General Scheme of Sidenor premises

The overall process for steel makers that handle scraps as raw material starts at the **scrap yard,** where the scrap material is selected for being loaded into the Electrical Arc Furnace. The incoming scrap is delivered by trucks and is sorted within dedicated boxes and areas at a scrap yard, according to the delivered scrap type's chemical composition. The scrap is mixed in a specific and optimized quantity to be loaded by cranes & baskets into the Electric Arc Furnace (EAF).

The general objective of the **melting process at the EAF** is to produce liquid steel with weights of about 130 t (at Sidenor plant) and narrow targets of chemical composition and temperature. The main parts of the electric arc furnace are the roof; which is provided with three holes through which the electrodes are inserted; and walls (both roof and walls cooled by water); the hearth that includes metal and slag and a tilting mechanism used to pour the molten metal to a ladle (tapping). For the specific case of Sidenor, the tap-to-tap cycle of the EAF is made up of several steps consisting of:

- Furnace charging: during the operation 2 baskets are loaded to produce one single batch of liquid steel (130 t) under an optimised mix and taking into account the scrap chemistry.
- Melting: The melting process starts right after the first scrap basket is loaded into the furnace. Once the scrap has been melted, the temperature is normally increased so that refining reactions can be carried out.
- Sampling: Temperature and chemistry are measured by Celox sampling method.
- Tapping: if temperature and chemistry are correct, the liquid steel is poured in the ladle that leaves the furnace to the next step in the steel shop (secondary metallurgy).

Similar protocols are used also in other melting shops. Although, the basic processes of the EAF appear quite straight forward consisting of simply providing enough electrical energy to heat and melt the steel scrap, chemical energy is to be supplied through three injection modules (oxygen and natural gas), which implies additional process data for AI modelling purposes.

After melting and refining in the EAF, the liquid steel is tapped into a ladle as mentioned, and alloy materials are added. The steel melt is further refined at a Ladle Furnace (LF) and a Vacuum Degassing (VD) station. Finally, the liquid steel melt is casted and solidified within a continuous casting plant to produce steel billets or blooms.

2.2.3 Current situation (i.e. operating modes)

Raw materials are the most expensive part in electric steelmaking with 70-90 % of the total production cost with energy consumption as the second largest cost (10-15 %). Therefore, an efficient use of raw materials and energy is of outmost importance in order to keep steel production at a competitive level, as errors in the calculations cause excessive usage of resources such as alloy elements and energy.

Moreover, the current driver towards zero emissions is forcing those steelmakers which manufactures steel by blast furnace to change to electric arc furnace. This kind of equipment are more efficient as they do not require a constant coke supply; instead, they use electricity carried through graphite electrodes to create an arc which minimizes the CO2 emissions. Thus, it is envisaged that the scrap will be a critical raw material.

This driver shows the mayor importance of new AI based solutions for improving the efficiency of all steel makers.

The operating mode of steel makers is characterized by uncertainty in the scrap properties. A specific example for Sidenor is explained hereinafter:

Around 5000 t/day are entering every day at Sidenor premises and more than 20 different scrap types are involved in the process for special steel manufacturing. The traceability of tramp elements is rather difficult and the optimum scrap to be charged to the EAF implies the usage of predictive models to calculate the optimum amount of scrap material that is required.

There are different models available for scrap supervision and mix optimization. As an example**:**

- **Scrap Supervision Model:** Developed under the Supercharge project frame. It consists of using advanced statistical methods for early detection of incorrect properties of the charge materials.
- **Scrap Mix Optimizer:** Developed in the OptiScrapManage project frame. It consists of using optimization tools for the adoption of the best mix to be loaded into the furnace to reduce the production cost.

However, the aforementioned models act as “stand alone” applications. Moreover, the models are based on excel files difficult to be updated and used in an industrial scale.

On the other hand, there are additional models for the EAF, like the **Dynamic EAF Model** developed in the Revamp project frame. This model consists of a dynamic predictive algorithm for the continuous calculation of the composition and temperature of the steel melt at the EAF. However, this model is rather difficult to be implemented in a steel shop, as many process data are required every 10 seconds, so that the output of the process can be predicted.

Within the s-X-AIPI project, a more autonomous and holistic approach is proposed for predicting the steel grade chemistry and temperature at tapping of the EAF depending on characteristics of the scraps loaded. Thus, especially attention has to be paid to some elements like Cu and P, trying to address the amount of these elements in the different scraps. This scrap supervision and further behaviour of the EAF process is proposed to be addressed under a machine learning approach.

To fulfil these requirements the proposed architecture has to react to anomalous or “unknown” situations which is considered of major importance.

2.2.3.1 Scrap characterization

The scrap supervision activity in steel shops relies on the Scrap Yards Operator’s experience who are in direct contact with Scrap Suppliers. Normally there is no complete chemical analysis of each scrap available, but some indications on the maximum limits of certain elements, shape and origin of the scraps. This implies that the loading of the furnace is based on some general inputs. In several occasions, the final chemistry is out of the desired range due to the punctual chemistry of certain scraps (uncertainty). In this detrimental scenario, the manufactured steel grade does not meet the customer requirements, which causes related losses of resources.

The s-X-AIPI approach will try to address the chemistry of the scraps based on a machine learning approach (multilinear regressions) together with an algorithm to assess the uncertainty.

2.2.3.2 Scrap mix optimization

The scrap mix optimization activity in steel shops is a duty to be accomplished by Scrap Yards Operators. The scrap to be loaded to the furnace is calculated based on the desired chemistry to be achieved and available scrap types. The calculation is normally based on simple mass balances.

There are other alternatives to calculate the scrap to be loaded based on optimization models, which normally are focused on minimizing the overall cost. However, this kind of models are rather difficult to be applied by operators due to the necessity of having an accurate input of the chemistry of the scraps.

Within the s-X-AIPI project an improved solution for optimizing the mix is proposed. The artificial intelligence to be developed will have as an input the accurate status on the scrap’s chemistry, which will be calculated by the ML solution explained in the previous section.

Thus, operators will end up having a tool which will address the optimal amount of material to be loaded to the furnace under minimization of overall costs and resources. Furthermore, this tool should be as autonomous as possible.

2.2.3.3 Steel temperature and chemistry

Temperature and chemistry within the electrical arc furnace are normally addressed by CELOX measurement at the ending of the melting process. Thus, the Operators of the Furnace account with the furnace status so that the electrical energy can be adjusted for ending in the desired scenario.

During the melting, some variables are monitored continuously like the electrical energy consumption, natural gas or the injected lime. This information will be used as an input of an algorithm based on machine learning (to be developed), so that the final status of the furnace (temperature and chemistry of some key elements) can be predicted.

Thus, operators will end up having a tool which will address the optimal poring moment and even to anticipate tapping moment.

Table 4. Operating current situation modes

Platform or system	Data availability	Involved users	Objectives	Performance indicators
Scrap characterization	Scrap types and amount of scrap loaded to the furnace.	Operators for the scrap yard	Calculate the chemistry of the different scrap types	Deviations of scrap average chemistry
Scrap mix optimization	Type of steel grades to be manufactured and scrap types to be used	Operators for the scrap yard	Suggest optimal scrap mix lowering the overall cost	Lower cost with accurate final chemistry
Steel temperature and chemistry	CELOX measurements at the end of process	Operators for the furnace	Calculate the ending process temperature and chemistry	Deviations in terms of chemistry and temperature

2.2.4 Future situation: improvements with self-X solutions

The steelmaking process in a future that is envisioned by the s-X-AIPI project still involves the operator taking decisions about machine set points, loading material and chemistry control. However, contrary to the current situation the developed support system will provide Operators with insights about the scrap properties and furnace temperature and chemistry (Figure 16).

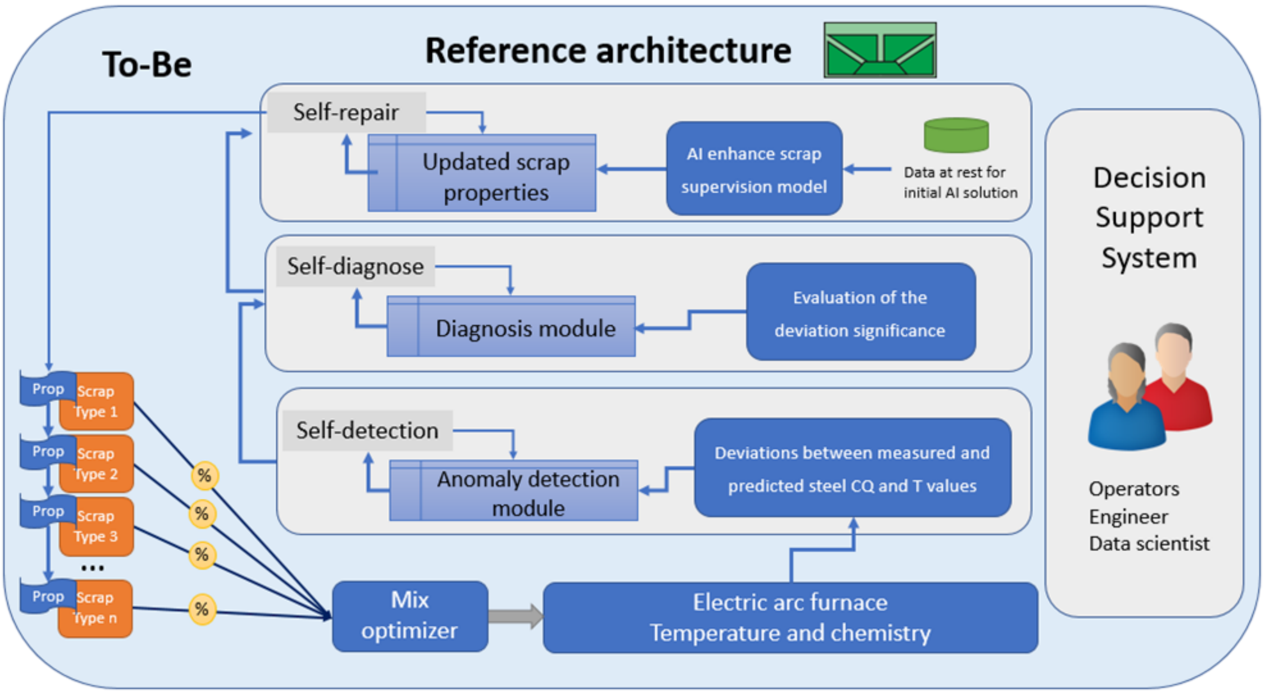


Figure 16. Steel use case autonomous manager

The envisioned self-X solution will monitor the input and output data space to detect unexpected events or drifting of the tools' accuracies and propose corrective actions to regain and maintain the robustness of the system. Overall, the goal is to ensure resilience and accuracy on the temperature and chemistry of steel at the end point of the furnace treatment.

2.2.4.1 Human roles interacting with self-X solutions

The goal of the system developed for the steel use case will be to facilitate the interaction of the operators, engineers and data scientist with the different modelling tools via a single interface to the autonomous manager (see Figure 16). The role of the human interacting with the self-X solution will be as follows:

- The enhance scrap supervision model will improve the available information on scrap properties which is key information for Scrap yard Operators. The accuracy of the model is envisaged to be high, therefore the feedback of Operators in the loop on the safety coefficients to be included into the code will be relevant.
- The knowledge about the prediction on the status of the chemistry and temperature at the ending of the furnace step will serve as a guideline for Furnace Operators. The feedback on the reliability of the prediction under operation will serve to improve the ML calculation.
- The autonomous manager will serve to data scientist to a studio of model independence. Thus, the autonomous manager will serve information about self-X-detection and diagnosis of anomalies during the process and give a suggestion of self-X-repair to data scientists.

The role of operators will be mainly to use the visual information as decision making support, while engineers or data scientist will have additional configuration possibilities.

2.2.5 KPIs: Baseline values and expected Results

IMPORTANT: Certain content is confidential, and it has been removed at the express request of SIDENOR, after approval of the PO.

KPI Category	Code from DoA	KPI definition	Critical-ity	Proposed for-mula/unit	*Baseline*	*Target value (%)*
Business KPIs	KPI1.1_Improvement of Productivity	**Productivity increase** The increase in productivity is proposed as a consequence of: • Decreased EAF Power on time, • Metal yield optimization and • Liquid steel quality improvement resulting in a de-creased rate of downstream quality rejections.	High	sum (tn of solid steel)/year	*confidential*[3]	(+5%)
Business KPIs	KPI1.2_Improvement of quality	**Liquid steel quality improvement** The improvement in liquid steel is proposed as a consequence of: • Better adjustment of residual elements to customer specifications, • Better internal rejections due to quality defects down-stream and • More controlled EAF process resulting in lower oxygen content and therefore lower impurities generated in SecMet process.	High	sum (tn of liq-uid steel)/year	*confidential*[3]	(+5%)
Technical KPIs	KPI1.3_Improvement of response time	**Decrease of Power on time in about 2 min** The decrease of power on is proposed as a consequence of: • A better prediction on liquid steel temperature, oxygen and phosphorous content would help the operators to better anticipate tapping moment and • A robust system for scrap characterization with super-vised feed to the optimizer will enable the possibility of updating scrap mix recipes more frequently, from monthly basis up to daily basis.	Medium	sum(min power on)/year	*confidential*[3]	(-4%)

[3] **IMPORTANT**: Certain content is confidential, and it has been removed at the express request of SIDENOR, after approval of the PO.

Technical KPIs	KPI4.3_Reduction in use of resources: scrap consumption	**Reduction of Scrap used** The decrease in scrap usage is proposed as a consequence of: • Optimizing the metal yield • Reducing quality rejections that leads to overall reduction of material resources	Medium	sum (tn of scrap)/year	*confidential* [3]	(-5%)
Technical KPIs	KPI4.4_Reduction in use of resources: electrical energy consumption	**Reduction of Electrical Energy** The decrease in the electrical energy used is proposed as a consequence of: • Direct correlation with the decrease of power on of 4%	Medium	sum(GWh)/year	*confidential* [3]	(-4%)
Environmental KPIs	KPI2.1_Reduction in generated CO2 emissions	**Reduction of CO2 emissions** The decrease in CO2 emissions is proposed as a consequence of: • Power on decrease in the EAF translates in specific energy consumption reduction	Medium	sum(electrical energy)*factor kgCO2eq/kWh / year	*confidential* [3]	(-4%)
Social KPIs	KPI5.3_Increase of competence, skills and qualification	**More attractive and safer jobs** • Scrap supervision tool with better understanding of scrap chemical composition for operators and the related increase of competence • Easy to use decision support system for operators and managers with explainable decision making	Medium	Survey to operators	N/A	4/5 acceptance

2.3 Pharma

2.3.1 Objectives/Benefits

The pharma use case deals with the chemical processing of suspensions during the preparation of active pharmaceutical ingredients or their intermediates. Processing of suspensions if often problematic, particularly during continuous operation, due to potential clogging issues in fluidic connections and piping or fouling of the reactor from particle agglomeration. This problem is of especial significance during electrochemical processing of suspensions, as ion concentration gradients and electrostatic interactions can provoke that solid materials accumulate on the electrode surface, ultimately leading to elevated cell potentials and electrode fouling and corrosion.

The main goals of this use case are:

- Develop automated, AI-based techniques for the real time monitoring of particles in suspension during chemical processing (in particular electrochemical processing) using optical coherence tomography (OCT). More specifically, monitoring of depositing of solid particles with electrode surfaces and potential electrode corrosion will be developed.
- Enable automated system response to negative interactions between the particles and the reactor, such as fouling or corrosion, incorporating s-X-AIPI for detection of unexpected outcomes.

It is expected that the s-X-AIPI project will significantly improve the reactor performance and thus decrease energy consumption during electrochemical processes. Early detection of reactor fouling or corrosion will maximize power efficiency during electrolytic reactions and ensure constant quality parameters for the output product, minimizing chemical waste.

2.3.2 General Description

The s-X-AIPI pharma case focuses on chemical processes involving solid suspensions in organic solvents. In particular, the electrochemical processing of cortisone leading to adrenosterone will be used as a case example. In a typical process, raw cortisone and a supporting electrolyte (e.g., sodium tetrafluoroborate) are suspended in a mixture of acetonitrile and water. The resulting suspension is pumped from the starting mixture tank to an electrochemical reactor, where the suspension is electrolyzed under a constant current of 3.3 mA/cm2. The reactor outlet is connected to the suspension reservoir, in a so-called flow electrolysis in recirculation mode (Figure 17).

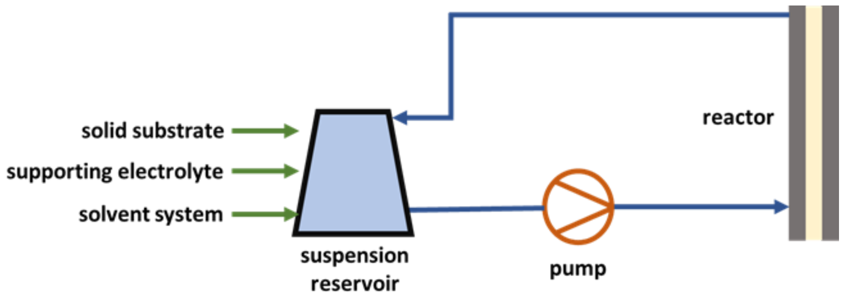


Figure 17. General scheme of an electrochemical process using a flow electrolysis cell in recirculation mode

During the chemical or electrochemical process, the particles in suspension may change in size or shape. In some cases they can gradually disappear (e.g., if the substrate is insoluble in the solvent system and the product is fully soluble), gradually appear (if the product is insoluble in the solvent system) or suddenly appear (precipitation). In any case, the presence of particles in suspension has a significant influence in the process. The rapid flow of particles in the system may cause issues in pumping systems, piping or the reactor. For an electrochemical process, additional issues may arise from electrode fouling by particle deposition or electrode corrosion. The latter case is not only relevant for the manufacturing of pharmaceutical ingredients, but also for other important fields such as flow batteries and fuel cells for energy storage. A suitable analytical tool for the real time monitoring of electrodes, as well as IT solutions for their control, is highly desirable.

2.3.3 Current situation (i.e. operating modes)

There is currently no operando monitoring techniques for early detection fouling or corrosion provoked by solid particles in a chemical or electrochemical reactor during a process. In a typical reaction, the raw materials are introduced in the system and the process initiated. The outcome of the chemical reactions is determined by off-line high performance liquid chromatography (HPLC) once the reaction is finalized. During the process, aliquot samples may be retrieved from the suspension reservoir and analysed by HPLC to monitor the reaction progress over time. Reaction process analytical technologies (PAT) are currently being developed to automatically obtain real-time data from chemical process via in-line, on-line or at-line HPLC, infrared (IR) or nuclear magnetic resonance (NMR) spectroscopy.

The effect of solid particles in suspension in a chemical process may be manifold. Clogging of the pump, system fluidics and connection or the reactor itself are rare, as adequate pumping systems and piping dimensions are selected according to the properties of the suspension. However, unexpected particle properties from a different batch of raw materials may cause clogging issues. Issues related with electrochemical processing of suspensions is less understood. Concentration gradients within a flow cell derived from the cell potential may cause rapid deposition of solid particles on the electrode surface, reducing its conductivity and thus increasing the cell potential and power consumption. Ultimately, corrosion of the electrodes may take place.

Consumption of electricity in order to facilitate the electrochemical reaction is by far the most significant use of power in a chemical plant. The consumption depends on the current used for the reaction, the resulting voltage as well as the process time. One goal of the self-X solutions is hence a reduction of the energy consumption. However, this must be done by also adhering to a reduction of input materials as well as preventing electrode damage. In the current situation, this is done by trial-and-error and best-effort approaches. The application of self-X solutions to the current use case aims at providing a systematic approach to facilitate this multi-objective optimization (Figure 18).

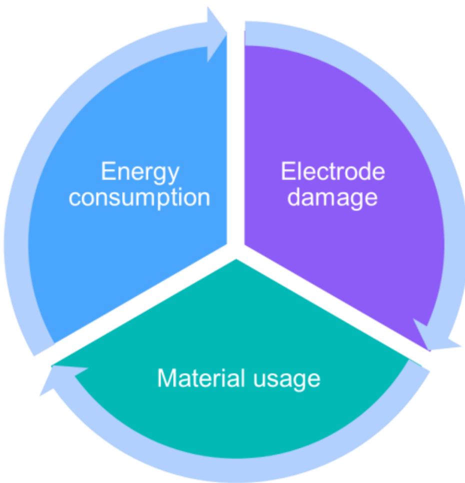


Figure 18: Parameters to optimize during the electrochemical reaction.

2.3.3.1 *Electrode monitoring*

The cleanliness of the electrodes in an electrochemical process is crucial for the successful completion of the process and a high conversion rate. Electrode fouling and corrosion can currently only be monitored by visual inspection. Small amounts of particles or corrosion in an early stage is however very difficult to see and will typically only be detected in more severe stages. Figure 19(a),(b) shows visual damage on a steel electrode and Figure 19(c),(d) shows a new and a corroded graphite electrode.

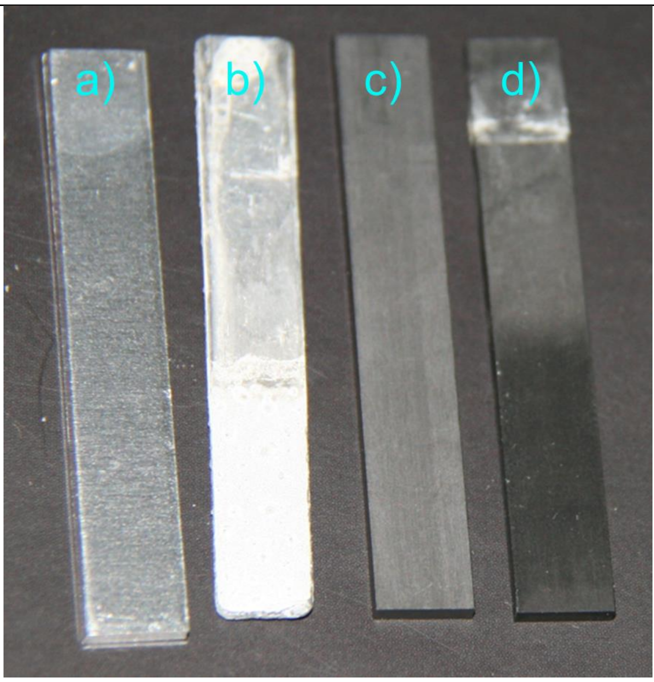
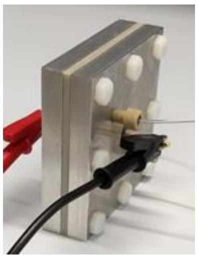
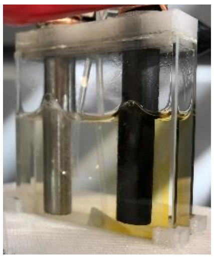


Figure 19: Comparison of (a) clean steel electrode, (b) damaged steel electrode, (c) new graphite electrode and (d) a corroded graphite electrode.

In quartz cells this can be done by looking through the glass of the cell. On the other hand, flow cells are designed for continuous synthesis and feature a closed reaction vessel. In the case of flow electrolysis cells, visual inspection may only be carried out by disassembling the reactor for examination of the parts.



(a) Flow cell



(b) Quartz glass cell

Figure 20: Two different types of cells used for an electrochemical process.

The late detection of corrosion typically damages the electrode beyond usability and leads to increased waste and energy consumption.

2.3.3.2 Process performance optimization

The amount of current required for a successful process varies with the chemistry and materials used. A current that is too low will not facilitate the reaction. On the other hand, if the current is too high corrosion of the electrode is likely. In the current situation, optimization is done manually by a trial-and-error approach. This leads to inefficient processes, a higher amount of waste as well as damaged electrodes due to improper current settings.

2.3.3.3 Current methods to identify process issues

Current methods to identify reactor issues related to the presence of solids in the reactor include:

- System pressure: an increase in the system pressure typically indicates clogging issues or buildup of material in the reactor walls. Pressure increase is usually detected at a late state.

- Cell potential: in the case of electrochemical processes, electrode fouling may be identified by an increase in the cell voltage. However, cell voltage increases may also be related to changes in the properties of the reaction mixture during the process (e.g., conductivity).

However, the aforementioned methods do not provide early detection of reactor issues.

2.3.4 Future situation: improvements with self-X solutions

The general objective of the self-X solutions in the use case is the enhance operation of the electrochemical process. Novel monitoring techniques such as OCT and the online evaluation of IR will provide an in-depth process understanding and form the basis for the development of an automatic optimization through a control strategy.

Combination of OCT technology, automated image analysis and ML-based control strategies will enable rapid detection of issues in the reactor and implement correction measures. The self-X algorithm will play a key role in ensuring appropriate adoption of the monitoring technology. OCT imaging is a sensitive technique, in which image quality is of high relevance. Moreover, other system failures (see below) might not be correctly identified by the ML-algorithm, thus requiring the intervention of the user.

In general, early detection of reactor anomalies such as deposition of solid material on the electrode surface or corrosion will permit rapid (automatic) correction in the process parameters. This will avoid (1) increase of the cell voltage, significantly reducing energy consumption, (2) increase the lifetime of the installed electrodes by reducing their corrosion and (3) reduced waste generation by minimizing out-of-spec materials obtained from the process.

Implementation of self-X solutions will ensure that human intervention is put in place when system failures cannot be amended by typical ML-based control strategies. This case includes malfunctions in the OCT optics or simpler failures such as electricity shortages or pump pressure oscillation.

Table 5. SELF-X capabilities overview

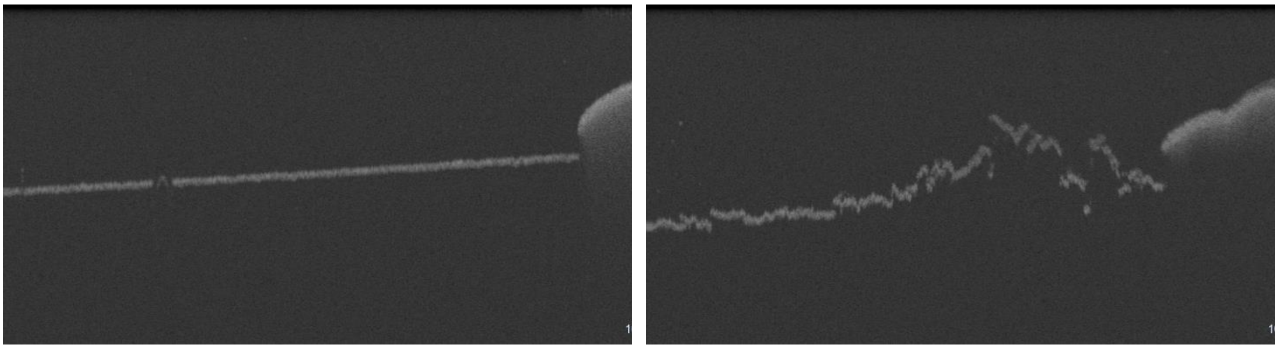
AM[1] capabilities	SELF-X detect	SELF-X diagnose	SELF-X repair	SELF-X optimization	Feedback for human in the loop
OCT image quality	Deviations in image quality	source of fault e.g. sensor problem, input changed…			detected problem and diagnosis
Process ML model predictions	if input and output data is within range	source of problem, e.g. input changed, predictions inaccurate…	Remove outliers and repeat training, report success of repair		problem and repair success
Control strategy	Information about reliability of input			Dynamically chooses optimal process settings	Human in control of ON/OFF

2.3.4.1 AM capabilities: Electrode monitoring via OCT

This self-X solution is designed to monitor the electrode during operation. As such, the operando monitoring of the electrode itself presents a novelty in electrochemistry. Crucial information about the electrode status can be obtained in real-time:

- Optical coherence tomography is used to scan an area on the surface of the electrode periodically
- Advanced real-time image analysis will provide fast and reliable feedback about the status
- Faults such as corrosion or fouling can be detected in very early stages, well before they could be seen visually

Figure 21 shows OCT images of an electrode in (a) the undamaged state and (b) after some corrosion occurred. It can be seen, that the corrosion can be noticed by the operator. The AM capabilities for the OCT images will include an automatic real-time evaluation of these images and a classification of the electrode status.



(a) **(b)**

Figure 21: OCT images of a (a) undamaged and (b) a corroded electrode.

2.3.4.2 AM capabilities: Process model

The automatic real-time information stemming from the OCT image analysis provides one input feature for the surrogate process model. Together with the information from IR and the power supply, a predictive model can be developed. This model will be capable of predicting the current concentration of converted material as well as a trajectory of the expected conversion rate for the remainder of the process.

With the details from the process model's predictions, the operator is provided with crucial information to make adjustments. These adjustments include for example lowering the current in case corrosion of the electrode is already visible.

Further, the role of the process model includes the prediction of corrosion or fouling. In the first case, the electrode gets damaged irrevocably. In the second case, the reaction is hindered but the electrode can be cleaned and hence re-used. The production of electrodes requires significant resources and hence the this AM is aimed at reducing the number of damaged electrodes during the experiments.

2.3.4.3 AM capabilities: Control strategy

Understanding the reaction dynamics in detail is the basis for an automatic control strategy in order to optimize the process. Several factors play a role for the success of the process:

- Percentage of material converted at the end of the process (conversion rate)
- Re-usability of the electrode after the process, i.e. no significant corrosion had occurred
- Optimization of energy used during the conversion

The automatic control strategy will be used to regulate the current with two objectives: (1) Damage on the electrode should be prevented and (2) the current should be high enough to allow for a smooth and fast reaction to take place.

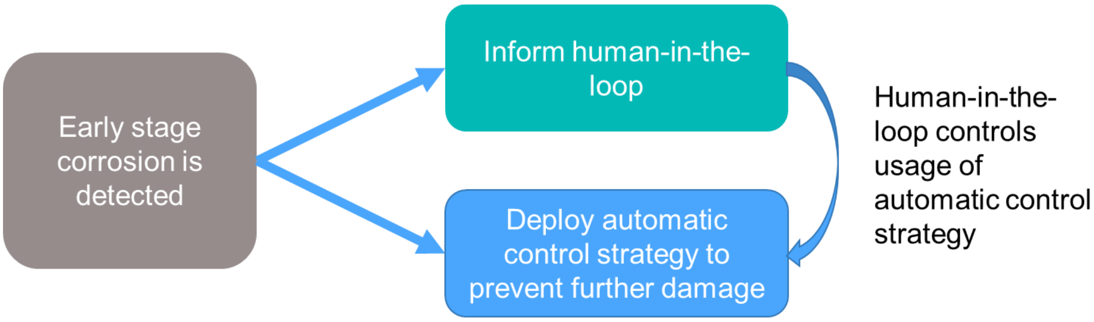


Figure 22: Reaction of the automatic control strategy upon early-stage corrosion.

2.3.4.4 Human roles interacting with self-X solutions

The self-X solutions shall identify the type of user required for the interaction depending on each specific case scenario. Technical issues related to equipment parts (e.g., OCT optics) will require intervention by equipment experts. Process related issues should be communicated to the reactor operator. Other decision-making processes will ultimately require the intervention or interaction of the plant manager.

There are several stages during the process where operator interaction is necessary or desired:

1. Manual adjustment of the position of the OCT probe
2. Activation or deactivation of the automatic control strategy
3. React upon notices of material faults.

The involvement of the human-in-the-loop for the specific self-X solutions is summarized in Table 6.

Table 6: Overview of human-in-the-loop involvement with the self-X solutions.

Self-X Pharma solution	Impact to human-in-the-loop	Feedback from human-in-the-loop
OCT images	The operator receives real-time information about the status of the electrode. The position of the OCT probe is evaluated and reported to the operator.	Design the automatic evaluation strategy used for the OCT images, classify the faults seen in the images used for algorithm training. The operator has to manually re-position the OCT probe if a dislocation is detected.
Process model	Provides real-time predictions about the expected conversion rate or expected electrode damage. Further, unexpected scenarios such as setup errors are highlighted.	The operator can react upon setup errors.
Control strategy	The current is automatically regulated in such a way that it does not damage the electrode but optimizes the process outcome.	In case problems are detected, the human-in-the-loop can switch the automatic control off.

2.3.5 KPIs: Baseline values and expected Results

KPI Category	Code from DoA	KPI definition	Criticality	Proposed formula/unit	*Baseline*	*Target value (%)*
Business KPIs	KPI1.5_New business model adopted	OCT for electrode supervision Extended process monitoring with self-X as advanced PAT solution	High	Presentation of self-X solution at relevant trade shows and conferences.	N/A	
Technical KPIs	KPI1.1_Improvement of Productivity	Optimization of the reaction parameters according to the input suspension properties will increase productivity by at least 30% via: less discarded out-of-spec materials (>10% production increase), increase of reaction efficiency (>15% production increase) and reduced process downtime for electrode exchange (>5% production increase)	Medium	Evaluation of time required for successful vs. unsuccessful processes.	N/A	
Technical KPIs	KPI1.2_Improvement of quality	The AI solution will help to reduce chemical waste and ensure constant quality attributes for the products: the pharmaceutical ingredients will have constant properties and failed drug batches that were previously discarded (diverted to waste) will be prevented. Prevention of anode corrosion will reduce waste further, as the life of the carbon material will be significantly extended. Autonomous computing will control the AI solution and establish reliable usage with human expert interaction.	Medium	Evaluation of material used in successful vs. unsuccessful processes and quantification of undesired side products (scrap). [kg of scrap per kg of API]	0.2	0.18 (-10%)
Technical KPIs	KPI1.3_Improvement of response-time	Reliable AI-assisted chemical development will provide faster process research and accelerate process scale-up. Self-X will enable immediate action in case of process troubles, thus reducing potential down-times. Manufacturing cycle (throughput) time by 20%	Low	Evaluation of time required for successful vs. unsuccessful processes	variable	(-20%)

Environmental KPIs	KPI2.1_Reduction in generated CO2 emissions	Electrode monitoring and corrosion prevention will avoid the generation of graphite waste from corroded electrodes. The life expectancy of graphite anodes in the process is approximately 5 batches, which should increase with the AI solution to at least 25. Due to the large carbon footprint of graphite manufacturing (25 equiv CO2), carbon emissions for the overall process will be significantly improved	High	Number of experiments with same electrode	5	25
Environmental KPIs	KPI4.1_Reduction in chemical waste generation	Prevention of discarded batches will account to at least 10% of the API material. Considering the typical process-mass intensity (PMI) of pharmaceutical processes (ranging between 10 to 100), this may lead to a reduction of waste of 100 kg to 1,000 kg per 100 kg of API produced.	Medium	Waste generated before purification according to process mass intensity [Kg of waste per kg of crude API]	13	11.7 (-10%)
Environmental KPIs	KPI4.4_Reduction in use of resources: energy and corresponding reduction in generated CO2 emissions	Energy savings during material processing due to electrolysis with higher current efficiency and corresponding CO2 footprint	High	Electricity required with self-X for the production of several successful batches vs. electricity required without self-X. [kWh per kg API / kg CO2 per kg API].	2 / 0.6	1.4 / 0.4 (-30%)
Social KPIs	KPI5.3_Increase of competence, skills and qualification	• Accelerate human learning based on transferring operators' knowledge and best practices. Tedious, routine-based manual work becomes obsolete, and safety and quality reach unprecedented high levels. Facilitate to the human being the treatment of great amount of information in a more effective, efficient, creative manner. Facilitate the creation of	High	Survey to operators	N/A	4/5 acceptance

		new self-X AI applications thanks to the building block of the s-X-AIPI open source toolset. • Consideration of human factors (user acceptance and experience). New AI applications based on s-X-AIPI toolset) that adapts and enriches heterogeneous skill levels.				

2.4 Aluminium

2.4.1 Objectives/Benefits

Since 1984, IDALSA is dedicated to the manufacture of aluminium ingots following raw material recycling procedures. Aligned with its compromise for a more sustainable manufacture, the recycling process engages the recycling and recovery of secondary materials, reducing the amount of by-products, waste, and polluting emissions.

Within self-X, the development and deployment of AI-based technologies are expected to optimize the recycling process by supporting plant operators in daily operational decisions with an Intelligent Decision Support System (IDSS). The IDSS will support operators at the plant following a data-driven approach by considering information from the complete value chain of the aluminium product i.e., suppliers, availability of raw materials, chemical analysis of the raw material and final aluminium billet, as well as relevant process-related information.

One of the key challenges of the IDSS will be supporting the decision-making process of the recipe of materials i.e., combination of raw materials, for the aluminium manufacture, which is currently performed by specialists in the plant and it is highly dependant on human expertise. The IDSS will be supported by the development of an aluminium passport that will include aluminium information through its complete lifecycle in order to provide manufacturers and users of the aluminium value chain standardized information about the quality, properties and origins of the aluminium products.

2.4.2 General Description

The aluminium recycling plant of IDALSA is located in Pradillo del Ebro (Zaragoza, Spain). The aluminium recycling process consists on several stages, as shown in Figure 23, which includes the reception and storage of raw materials, dross classification and grinding (optional), primary melting, secondary melting and alloying, moulding and cooling, and the preparation of aluminium billets.

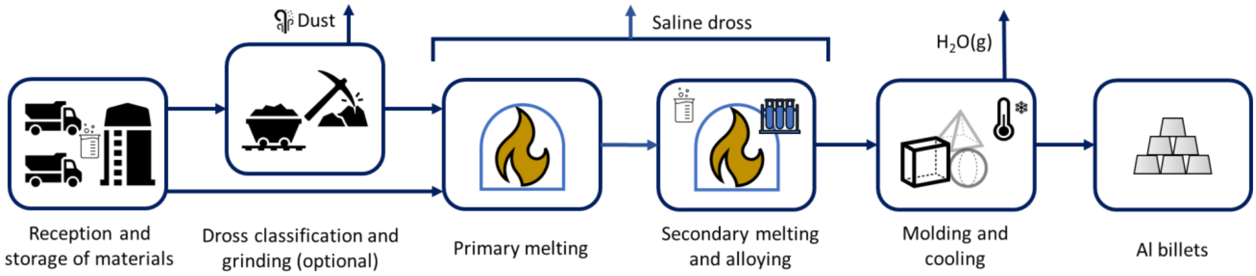


Figure 23. Process stages of IDALSA aluminium recycling process

The process starts with the reception of raw materials from different sources in trucks (AI waste collectors, industries, household waste, etc.). A chemical analysis is performed at the reception of materials to verify their chemical composition, prior to its sorting and storage. In case that the materials need further processing, they can be sent to the grinding facility to reduce its volume. Subsequently, according to the specifications of clients and plant-related information, materials are selected in the so-called recipe to be processed via primary and secondary melting and alloying, and the resulting recycled AI mixture is cooled and poured into moulds to be provided as billets, hemispheres, or pyramid-like formats. The main industrial facility in IDALSA consists of a factory line embracing two furnaces for primary melting, three furnaces for secondary melting and alloying, a line for cooling and moulding the aluminium mixture, and a separate lab to perform in-line quality control assessments, as depicted in Figure 24.

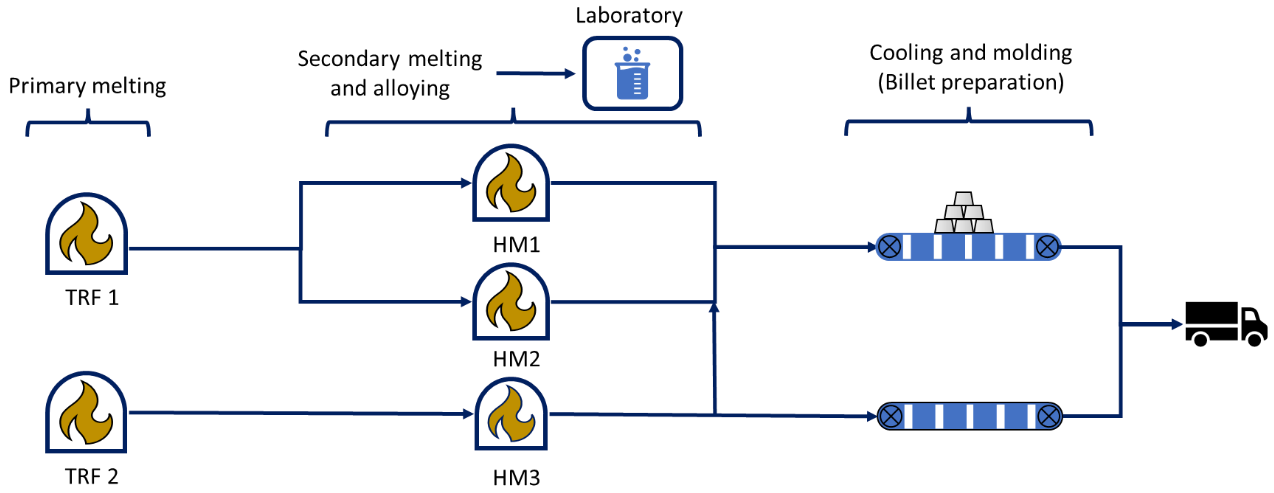


Figure 24. Configuration of IDALSA's main industrial facility

All process-related information is stored in a proprietary LINUX-based application (see Figure 25), including i) availability, chemical composition, silo location, and properties of raw materials, ii) aluminium recipes, and iii) clients and orders, among others. Besides that, chemical analysis, especially those analysis conducted during the alloying process, are stored locally in a PC at the main industrial facility. More information about these two systems is provided in Section 4.4 Aluminium.

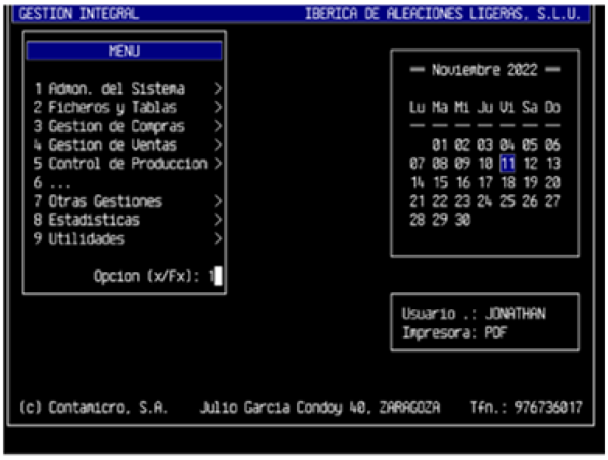


Figure 25. Screenshot of IDALSA LINUX-based application that stores all process information

2.4.3 Current situation (i.e. operating modes)

As previously introduced, the aluminium recycling process starts with the reception and storage of raw materials by different suppliers. At this stage, the incoming materials are weighted, and a chemical analysis is performed to evaluate their properties for their proper handling and storage. Currently, IDALSA counts with approximately 100 silos (exemplary photos are shown in Figure 26) for the different raw materials, which are stored according to their chemical properties, the ratio of solid/dust, and their humidity content. If needed, an intermediate step to grind raw materials is conducted to prepare the material for its processing, with the resulting extraction of dust that is ultimately detrimental for the performance of the process.

Figure 26. Exemplary silos at IDALSA with different raw materials

Based on the requirements and specifications of clients, referred as norms, and the overall business plan, the recipe of materials for the aluminium production is selected. This decision relies on the chemical composition of each raw material separately, the final chemical composition expected for the aluminium product, the cost of materials, their availability at the plant, and the silo where the materials are located, among others. To date, the ultimate decision is based on the expertise of specialists in the plant, and there does not exist any integrated software supporting this process.

Once the type and quantity of materials are selected, materials are brought from their silos to the main plant, where they are weighted with mechanic shovels and introduced in the charger of raw materials together with the melting mix. Materials are introduced in the rotative furnace (Figure 27, left) to be melted by oxygas combustion in a sequential manner, a process which is controlled based on the expertise of a specialist through visual evaluation of the melting process e.g., behaviour of the melting mass, colour, aspect, etc. The primary melting process lasts around 4 hours, before the melted aluminium is transferred to a second furnace for its secondary melting and alloying, a process that lasts approximately 20 minutes.

The melted aluminium is deposited by gravity in a smaller furnace (Figure 27, right) to maintain the liquid state of the aluminium mixture preserving completely mixed conditions. At this stage, the mixture is analysed by means of mass spectrometry to evaluate its composition prior to alloying. Based on the information from the chemical analysis and the specifications of the client, alloys are added in the aluminium mixture e.g., Cu, Si, Mn, Fe, etc. The process lasts between 2 and 3 hours and the chemical analysis is repeated to verify the properties of the melted aluminium before proceeding to the subsequent stage with the final aluminium mixture.

Figure 27. Left, furnace for primary melting; right, furnace for secondary melting and alloying

Once the melted aluminium meets the specifications established by the norm, it is transferred to moulds with a volumetric dispenser to be cooled by dripping. When the aluminium reaches a temperature below 600ºC, it solidifies according to the shape of the mould (Figure 28, left). The aluminium products are then unmoulded and piled alternatively in cubic blocks of 60x60x60cm ready for packaging and transport (Figure 28, right).

Figure 28. Left, aluminium moulds; right, aluminium billets

The main by-product generated by secondary aluminium melting is saline dross. The aluminium saline dross contains 15-30% aluminium oxide, 30-55% sodium chloride, 15-30% potassium chloride, 5-7% metallic aluminium and remaining impurities depending on their source i.e., carbides, nitrides, sulphides, and phosphides. After aluminium refinement, the total waste dross left in the environment as pollution can sum up to 3 million tons.

The IDSS in the aluminium use-case will support workers in their daily operational decision-making processes, more specifically, in the aluminium recipe i.e., selection of scraps and alloys to be used for producing the aluminium product. Table 7 presents the data, users, objectives and performance indicators of the proposed IDSS.

Table 7.Operating modes in the aluminium use case

Platform or system	Data availability	Involved users	Objectives	Performance indicators
IDSS - Aluminium recipe (mix)	Raw materials (chemical analysis, properties, silo location, availability, price), historical aluminium recipes (scraps and alloys), composition of melted aluminium mixture (before and during alloying) and end product.	Aluminium plant operators and managers.	Support in the selection of aluminium recipes (predictive algorithms) to optimize the use of resources ensuring compliance with product norms	- Deviations between the expected composition of the aluminium mixture and the real composition obtained. - Deviations of scrap properties (estimated vs. real)

2.4.4 Future situation: improvements with self-X solutions

Based on the challenges brought by the current situation, and specifically on the human-dependence level in the aluminium-making process in IDALSA, an Intelligent Decision Support System (IDSS) will be developed in self-X with the aim of supporting operators in their daily tasks focusing on the recipe selection of raw materials.

The IDSS will be originally feed with historical information considering information about raw materials (chemical analysis, silo location, availability at the plant...), previous recipes, and final chemical analysis of aluminium product, to develop an RL framework that would assist operators selecting future recipes. The proposed solution will optimize the selection process of raw materials with the aim of reducing waste, energy expenses, and emissions, while maintaining high-quality standards for the process and aluminium product.

The support in the recipe selection is envisaged to be two-folded as depicted in Figure 29. On one side, a general planning of recipes will be provided to the general manager in a periodic basis following current practices; on the other hand, online adaptive planning support will be provided with the aim of improving short-term planned heats based on the most recent information to tackle the uncertainty that might come, from instance, from the common storage of raw materials at silos or the initial chemical analysis that is performed when a new material enters the facility that might be unrepresentative of the complete material lot. Some of the supporting tasks of the IDSS considering the online adaptive planning support might include evaluating the general recipe and tackling issues brought by changing scenarios e.g., availability of new materials in the plant or deviations in the process, as well as supporting process operators in the alloying process to select the optimal alloys based on the chemical analysis performed after primary melting and the general recipe used.

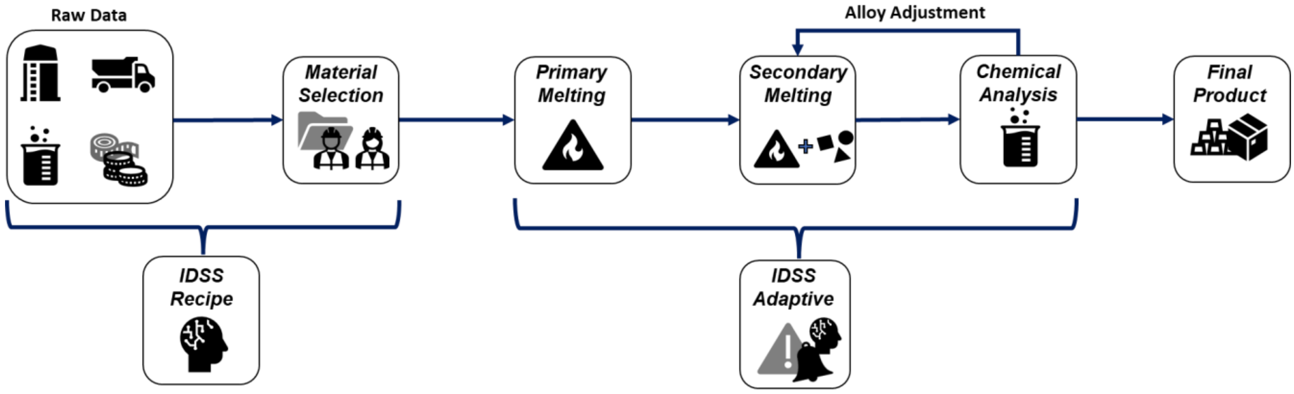


Figure 29. IDSS contribution in the process flow

Based on the proposed solution, Table 8 presents the self-X abilities that will be addressed by the IDSS.

Table 8. self-X abilities in the aluminium use-case

AM[1] capabilities	SELF-X detect	SELF-X diagnose	SELF-X repair	Feedback for human in the loop	Contributions
IDSS - Aluminium recipe	Deviations in scraps and materials used	Aluminium recipe with respect to the proposed recipe		Aluminium recipe suggestions, and expected composition to be obtained	Process, Circularity

The development of the RL-based IDSS will be accompanied by the generation of synthetic data using Generative Adversarial Networks (GANs) trained with the given historical data to augment the dataset. This generated data will have a distribution as close as possible to the original data. In this way, the IDSS will not need to rely exclusively in historical data, which might be limited, incomplete, or difficult to extract in an appropriate manner for AI developments.

In parallel to the discussed solutions, a product passport of the aluminium product will be developed in order to provide internal and external stakeholders of the aluminium value chain standardized information from the aluminium products covering its lifecycle i.e., information about their source, composition of materials, chemical analysis, etc.

As previously introduced, all process related information is stored in a proprietary LINUX-based application, while information from intermediate chemical analysis conducted during alloying are stored locally in a server directly from a PC located in the lab at the main industrial facility. More information about these two components can be found in Section 4.4. Aluminium.

2.4.4.1 Human roles interacting with self-X solutions

The proposed IDSS aims at supporting operators at the plant in their daily life tasks, working closely together with humans. Following this approach, the interaction between the IDSS and the operators will be dual and reciprocal (Figure 30). On one hand, the IDSS will support the recipe selection process based on previous experiences and historical and synthetic data to support operators in an user-friendly approach. The IDSS will provide periodic reports to the general manager with the suggested recipes that will include information about the materials selected, their silo location, overall cost expected, etc. Besides that, as part of the adaptive online planning support, specific reports will be provided to the general manager and plant operators for the alloying recipe to ensure that all products meet their norms at the end of the process.

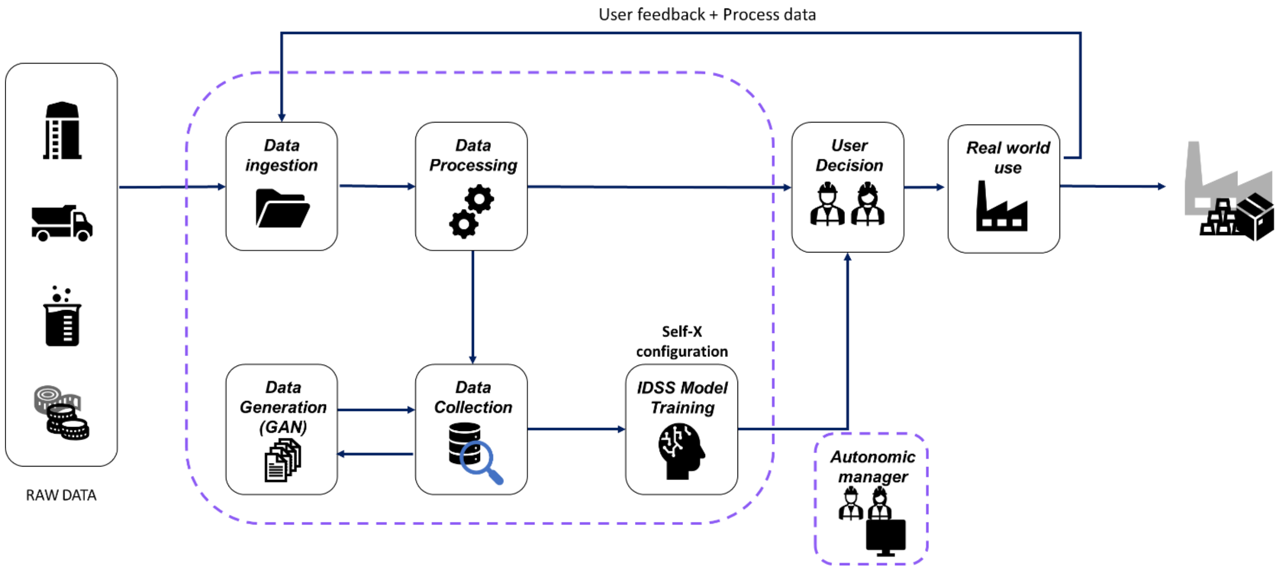


Figure 30. Integration of user feedback in the IDSS solution

On the other hand, as part of the dual experience expected in the IDSS, the operators and users will be expected to give their feedback on the solutions proposed by the IDSS, with the aim of improving its performance and enabling its future adaption to the skills or profiles of the operators, when possible. An in-depth evaluation to assess human acceptance of the proposed AI solutions will be carried out through their development in order to facilitate their use and integrate the feedback of operators during the fine-tuning of the AI models while reducing human workload on that task as much as possible i.e., properly analysing the essential information needed to optimize the AI models to avoid requiring non-valuable information. The feedback from operators will be used as corrective labels and/or new data for re-training the models as well as to optimize the exchange of information between the IDSS and operators. The impact of the proposed self-X solution to the human and the expected feedback are summarized in Table 9.

Table 9. Human in the Loop in the Aluminium use case

Self-X aluminium solution	Impact to human in the loop	Feedback from human in the loop
IDSS - Aluminium recipe	Enhanced knowledge about the aluminium production process e.g., expected composition of aluminium recipes, behaviour of materials, etc.	Evaluate the recipes suggestions (provide corrections or adjustments, if needed) and provide relevant information from the process performance

2.4.5 KPIs: Baseline values and expected Results

KPI Category	Code from DoA	KPI definition	Criticality	Proposed formula/unit	*Baseline*	*Target value (%)*
Business KPIs	KPI1.1_Improvement of Productivity	Productivity increase of 6% of total 30,000 t/year by increasing metal yield of aluminium scrap	High	tn Al/Y	30000	31800 (+6%)
Business KPIs	KPI1.1_Improvement of Productivity	Reduce costs (flux €/Kg-day) by 3 %	High	flux €/Kg-day	7687	7456 (-3%)
Business KPIs	KPI1.1_Improvement of Productivity	Reduce overall gas consumption (€/m3-day) by 4,50%	High	€/m3-day	4705	4494 (-4.5%)
Technical KPIs	KPI1.2_Improvement of quality	Increase the efficiency (Kg Al. prod/Kg Al. Theoric) by 3%	Medium	Al. fed (m3/Kg)	60%	61.8% (+3%)
Technical KPIs	KPI1.2_Improvement of quality	Decreased flux material: NaCl/Al by 7%	Low	tn NaCl/tn Al	4403	4095 (-7%)
Technical KPIs	KPI1.2_Improvement of quality	Decreased flux material: KCl/Al by 6%	Low	tn KaCl/tn Al	1887	1775 (-6%)
Technical KPIs	KPI4.4_Reduction in use of resources: energy	Reduce the gas consumption/material Al. fed 8%	Medium	gas consumption/material Al. Fed	1184	1089 (-8%)
Environmental KPIs	KPI4.3_Reduction in use of resources: saline dross	Reduce waste (e.g. saline dross)/by-products (Kg. salt slags/Kg. Al. fed) by 4 %	Medium	Kg. salt slags/Kg. Al. Fed	25300	24288 (-4%)
Social KPIs	KPI5.3_Increase of competence, skills and qualification	• Improve human knowledge about the process in a data-driven approach • Improve operators' wellbeing at work (relieving human's dependence in decision-making processes, improving operators' satisfaction at work) • Improve and support user training and transfer learning capabilities to new operators (incorporate knowledge gained in training programs and support novel operators in decision-making)	High	Survey to operators	N/A	4/5 acceptance

Social KPIs	KPI6.6_Artificial intelligence, machine learning	• User acceptability of the tool (measured by user engagement with the tool i.e., number of times the tool is used/called by operators) • Trustworthiness of the tool's suggestions (number of times recipes were incorporated into the process by operators)	Medium	Counter	N/A	Expected 85%

3 Requirement's Specification of the self-X AI solutions

This section is focussed to describe the requirement's specifications of each self-X AI solutions within the use cases described in previous section. First of all, 'User stories' are used to describe the requirements from an end-user perspective, and then requirements from each self-X AI solutions and system requirements are showed, divided in functional (F) and non-functional (NF) requirements.

As requirements references throughout all the uses cases the following unique ID has been used:

<Pilot>-<US/R/SR><self-X>.<Ascending id>

where:

- First <Pilot> is codified using AS for "Asphalt", ST for "Steel", AL for "Aluminium" and PH for "Pharma"
- Then <US/sXR/SR> is used to indicate "User Story", "self-X Requirements" and "System Requirements" respectively
- Next the different self-X solutions are numbered
- And finally, the different items are numerically coded starting with 01

3.1 Asphalt

Within Asphalt use case the requirements mapping is as follow:

User stories	self-X Requirements	System Requirement
AS-US1.1	AS-sXR1.01 AS-sXR1.02 AS-sXR1.08 AS-sXR1.09	AS-SR1.1
AS-US1.2	AS-sXR1.01 AS-sXR1.03 AS-sXR1.08 AS-sXR1.09	AS-SR1.1
AS-US1.3	AS-sXR1.01 AS-sXR1.05 AS-sXR1.06 AS-sXR1.09	AS-SR1.1
AS-US1.4	AS-sXR1.05 AS-sXR1.06 AS-sXR1.07	AS-SR1.1
AS-US2.1	AS-sXR2.08 AS-sXR2.09 AS-sXR2.10 AS-sXR2.11 AS-sXR2.12 AS-sXR2.13	AS-SR2.1

AS-US2.2	AS-sXR2.01 AS-sXR2.02 AS-sXR2.03 AS-sXR2.04 AS-sXR2.05 AS-sXR2.06 AS-sXR2.07	AS-SR2.1
AS-US2.3	AS-sXR2.01 AS-sXR2.02 AS-sXR2.03 AS-sXR2.04 AS-sXR2.05 AS-sXR2.06 AS-sXR2.07 AS-sXR2.08 AS-sXR2.09 AS-sXR2.10 AS-sXR2.11 AS-sXR2.12 AS-sXR2.13	AS-SR2.1
AS-US3.1	AS-sXR3.01 AS-sXR3.02 AS-sXR3.06 AS-sXR3.07	AS-SR3.1
AS-US3.2	AS-sXR3.03 AS-sXR3.04 AS-sXR3.05	AS-SR3.1
AS-US4.1	AS-sXR4.01 AS-sXR4.02 AS-sXR4.09 AS-sXR4.10	AS-SR4.2
AS-US4.2	AS-sXR4.01 AS-sXR4.03 AS-sXR4.09 AS-sXR4.10	AS-SR4.1
AS-US4.3	AS-sXR4.04 AS-sXR4.05	AS-SR4.1

	AS-sXR4.07	
AS-US4.4	AS-sXR4.04 AS-sXR4.06 AS-sXR4.08	AS-SR4.2

3.1.1 self-X #1 solution: "Asphalt mix design"

User stories "Asphalt mix design"	
AS-US1.1	As a manufacturing operator in an asphalt plant (foremen and engineers), I want to know the deviations between real and theoretical mixing times so that I can detect time deviations in the process that may influence the final quality of the mix and operational or logistic delays
AS-US1.2	As a manufacturing operator in an asphalt plant (foremen and engineers), I want to know the deviations between real and nominal cold aggregates feeder mass balances so that I can detect a malfunctioning of the cold aggregates bin Variable Frequency Drives (VFD)
AS-US1.3	As a manufacturing operator in an asphalt plant (foremen and engineers), I want to know the deviations between real and theoretical asphalt mix composition so that I can detect composition deviations in the process that may influence the final quality of the mix
AS-US1.4	As a data scientist, I want to be able to train classification model in a simple way so that the time I need to perform a retraining will be minimized

Requirements "Asphalt mix design"	F	NF
AS-sXR1.01 The self-X AI solution must import production data from "Connected Plant" system to an external data repository system (outside EIFFAGE infrastructure) where the data will be stored (a from which it could be queried)	X	
AS-sXR1.02 The self-X AI solution must allow consulting and comparing historical data related to real VS theoretical mixing times	X	
AS-sXR1.03 The self-X AI solution must allow consulting and comparing historical data related to real VS nominal cold aggregates feeder mass balances	X	
AS-sXR1.04 The self-X AI solution must allow consulting and comparing historical data related to real VS nominal asphalt mix composition	X	
AS-sXR1.05 The self-X AI solution must allow to a data scientist measure/check data quality as new data amounts to 5% of the total data used for training is received		X
AS-sXR1.06 The self-X AI solution must allow to a data scientist select the best classification models minimizing his/her participation in the process to only two choices: parameterizing initially and selecting a new model at the end		X
AS-sXR1.07 The self-X AI solution must allow to a data scientist measure/determine quality of classification model output when new data amounts to 5% of the total data used for training becomes available		X
AS-sXR1.08 The asphalt plant staff must be able to interact with the self-X AI solution through a web accessible dashboard	X	
AS-sXR1.09 Only registered users are able to use the dashboard	X	

3.1.2 self-X #2 solution: "Plant elements diagnosis"

User stories "Plant elements diagnosis"	
AS-US2.1	As a manufacturing operator in an asphalt plant, I want to know the minimum consumption of electrical energy and the minimum amount of fuel to be able to operate the plant efficiently
AS-US2.2	As an asphalt plant manufacturing operator, I want to know if any machinery is malfunctioning so that I can take corrective action
AS-US2.3	As a maintenance technician at an asphalt plant, I want to know the status of the plant's machines in order to schedule machine maintenance before a breakdown occurs

Requirements "Plant elements diagnosis: Predictive maintenance"	F	NF
AS-sXR2.01. The self-X AI solution must detect anomalies in the production variables (temperature, vibration, electrical consumption).	X	
AS-sXR2.02. The self-X AI solution should analyse the variables on demand and when a deviation is detected.		X
AS-sXR2.03. The Self-X AI solution must have a future horizon, the performance of predictive maintenance.		X
AS-sXR2.04. The Self-X AI solution will be able to suggest parameter adjustments and recovery actions.	X	
AS-sXR2.05. The Self-X AI solution will allow optimizing the production process and increasing production quantity.	X	
AS-sXR2.06. The Self-X AI solution will allow stabilizing y enhance production continuity, avoiding failures.	X	
AS-sXR2.07. The Self-X AI solution will allow the final decision to be triggered by humans through the user interface.		X

Requirements "Plant elements diagnosis: Burner efficiency diagnosis"	F	NF
AS-sXR2.08. The Self-X AI solution must detect anomalies in fuel consumption or temperature variation and perform a data scan to assess burner efficiency.	X	
AS-sXR2.09. The Self-X AI solution must be activated on demand or when a deviation is detected.		X
AS-sXR2.10. The Self-X AI solution must have as a horizon in the future the optimization of the operating parameters of the burner.		X
AS-sXR2.11. The Self-X AI solution will be able to suggest parameter adjustments and recovery actions.	X	
AS-sXR2.12. The Self-X AI solution will allow stabilizing y enhance production	X	

continuity, avoiding failures.		
AS-sXR2.13. The Self-X AI solution will allow the final decision to be triggered by humans through the user interface.		X

3.1.3 self-X #3 solution: "Paving conditions and parameters"

User stories "Paving conditions and parameters"	
AS-US3.1	As a manufacturing operator in an asphalt plant (foremen and engineers), I want to know in advance the temperature at which the asphalt mix will arrive at the job site so that I can modify current asphalt mix design and/or the manufacturing process to adapt it to the needs of the paving operation
AS-US3.2	As a data scientist, I want to be able to train predictive model in a simple way so that the time I need to perform a retraining will be minimized

Requirements "Paving conditions and parameters"	F	NF
AS-sXR3.01 The self-X AI solution must import logistic data from "IT ASPHALT" system to an external data repository system (outside EIFFAGE infrastructure) where the data will be stored (a from which it could be queried)	X	
AS-sXR3.02 The self-X AI solution must evaluate the relationships between the production data (temperature of the material at the plant outlet) and the laying and compacting data (temperature of the material on site)	X	
AS-sXR3.03 The self-X AI solution must allow to a data scientist measure/check data quality as new data amounts to 5% of the total data used for training is received		X
AS-sXR3.04 The self-X AI solution must allow to a data scientist select the best prediction models minimizing his/her participation in the process to only two choices: parameterizing initially and selecting a new model at the end		X
AS-sXR3.05 The self-X AI solution must allow to a data scientist measure/determine quality of predictive model output when new data amounts to 5% of the total data used for training becomes available		X
AS-sXR3.06 The manufacturing operator in an asphalt plant must be able to interact with the self-X AI solution through a web accessible dashboard	X	
AS-sXR3.07 Only registered users are able to use the dashboard	X	

3.1.4 self-X #4 solution: "Quality traceability at lab level"

User stories "Quality traceability at lab level"	
AS-US4.1	As a manufacturing operator in an asphalt plant (foremen and engineers), I want to verify that the mixture that is being produced complies with the specifications given by the laboratory and with the standards so that I can adapt the asphalt mix design to improve its quality

AS-US4.2	As a laboratory responsible in an asphalt plant, I want to obtain predictions about the volumetric/mechanical properties of an asphalt mix so that I will anticipate possible anomalies in the mix to manufacturing operators
AS-US4.3	As a data scientist, I want to be able to train predictive model in a simple way so that the time I need to perform a retraining will be minimized
AS-US4.4	As a data scientist, I want to be able to train classification model in a simple way so that the time I need to perform a retraining will be minimized

Requirements "Quality traceability at lab level"	F	NF
AS-sXR4.01 The self-X AI solution must import laboratory data from "HCLAB" system to an external data repository system (outside EIFFAGE infrastructure) where the data will be stored (a from which it could be queried)	X	
AS-sXR4.02 The self-X AI solution must be able to detect gradation deviations between asphalt mix design and designed formula	X	
AS-sXR4.03 The self-X AI solution must include predictive models developed to predict the volumetric/mechanical properties of an asphalt mix.	X	
AS-sXR4.04 The self-X AI solution must allow to a data scientist measure/check data quality as new data amounts to 5% of the total data used for training is received		X
AS-sXR4.05 The self-X AI solution must allow to a data scientist re-configure the predictive models minimizing his/her involvement in the process to only two choices: parameterizing initially and selecting a new model at the end		X
AS-sXR4.06 The self-X AI solution must allow to a data scientist select the best classification models minimizing his/her participation in the process to only two choices: parameterizing initially and selecting a new model at the end		X
AS-sXR4.07 The self-X AI solution must allow to a data scientist measure/determine quality of predictive model output when new data amounts to 5% of the total data used for training becomes available		X
AS-sXR4.08 The self-X AI solution must allow to a data scientist measure/determine quality of classification model output when new data amounts to 5% of the total data used for training becomes available		X
AS-sXR4.09 The asphalt plant staff must be able to interact with the self-X AI solution through a web accessible dashboard	X	
AS-sXR4.10 Only registered users are able to use the dashboard	X	

3.1.5 self-X abilities in asphalt use case prioritization.

Due to the quality and quantity of the historical data available to develop the models, solutions s-X#1 and s-X#4, Asphalt Mix Design and Quality, are prioritized, thanks to the fact that there is more data from the App Connected Plant and from the laboratory software, HCLab. It is planned to start in these solutions in M9.

Moreover, thanks to the existing synergies with the s-X#3 and s-X#4 solutions, Logistic and Laboratory, where auto-X procedures for predictive models will be defined, these solutions will be developed in parallel starting in M12. Thus, this challenge will be addressed contemplating the possibility of unifying solutions (not only in this use case, but also in all others) and taking advantage of what has been learned in each one of them.

3.1.6 System Requirement

Table 10. SR Asphalt mix design. self-X #1

<table>
<tr><td colspan="3">ID</td><td>AS-SR1.1</td></tr>
<tr><td colspan="3">Overall Description</td><td>Asphalt Mix components and key process parameters prediction model</td></tr>
<tr><td rowspan="8">Specific Requirements</td><td rowspan="2">Feature</td><td>Introduction & Purpose of feature</td><td>Considering available production data all pre-processing activities will be carried out and several clustering algorithms applied in order to classify (in an unsupervised way) all the asphalt mix design that are currently being produced.
Moreover, an intelligent dashboard will be developed in order to show to the plant operators those variables that are consider the most important ones.</td></tr>
<tr><td>Functional Requirements</td><td>Self-X Data Transform & Self-X Data Exploration: data coming from production
Self-X Model Training: Composition deviation, process parameters deviation and possible causes at production level</td></tr>
<tr><td rowspan="4">External Interface Requirements</td><td>User Interfaces</td><td>Intelligent dashboards to present the insights</td></tr>
<tr><td>Hardware Interfaces</td><td>Not apply</td></tr>
<tr><td>Software Interfaces</td><td>App Connected Plant (Production Data)</td></tr>
<tr><td>Communications Interfaces</td><td>Production data: MQTT</td></tr>
<tr><td colspan="2">Performance Requirements</td><td>Production data:
- data reception rate of 800Mb/month
- time series & alphanumeric data</td></tr>
<tr><td colspan="2">Other non-functional requirements</td><td>Not defined</td></tr>
</table>

Table 11. SR Asphalt mix plant and elements diagnosis. self-X #2

<table>
<tr><td colspan="3">ID</td><td>AS-SR2.1</td></tr>
<tr><td colspan="3">Overall Description</td><td>Maintenance prediction model & Maintenance prevention model</td></tr>
<tr><td>Specific Requirements</td><td>Feature</td><td>Introduction & Purpose of feature</td><td>• Improve the continuum of the process (avoid production shutdowns as much as possible)
• Improve maintenance efficiency (cost and time reduction)</td></tr>
</table>

			• Improve situational awareness of the motors health (useful for maintenance purpose but also for the process/product quality perspective)
		Functional Requirements	Data Exploration / Data Quality • Data from different motors over the manufacturing process (e.g. Dryer motor, Auger motor, Lifting motor, Fan motor, Introducer tape engine, Compressor motor) and their failures • Data for motors´ health electricity consumption • Data from Production (motor power, temperature, fuel consumption, etc.)
	External Interface Requirements	User Interfaces	
		Hardware Interfaces	Not defined
		Software Interfaces	
		Communications Interfaces	Maintenance manager
	Performance Requirements		Motors and burner efficiency and/or predictive maintenance and possible causes
	Other non-functional requirements		

Table 12. SR Paving conditions and parameters. self-X #3

ID			*AS-SR3.1*
Overall Description			Laying and compacting temperature prediction model
Specific Requirements	Feature	Introduction & Purpose of feature	Based on the temperature of the asphalt mix when it is loaded on trucks, as well as the conditions during the transport (mainly time and location), a prediction model will be generated trying to know in advance the temperature of the mixture during its laying and compacting on the road. Based on these predictions a recommender system could be generated, in order to modify the asphalt mix design. In that case that not so much historical data are available, so this model will be developed by combining expert rule-based systems with information extracted from data.
		Functional Requirements	Data Exploration/Data Quality: data coming from paving and logistic Self-X Model Training: Laying and compacting temperature model
	External Interface	User Interfaces	Intelligent dashboard to propose improvement within the asphalt mix design

	Requirements	Hardware Interfaces	Not apply
		Software Interfaces	Production data: App Connected Plant Paving and Logistic data: IT ASPHALT
		Communications Interfaces	Production data: MQTT Logistic data: SQL Server database, Mongo DB
	Performance Requirements		Production data: - data reception rate of 800Mb/month - time series & alphanumeric data Logistic data: - data reception rate of 36Mb/year - time series data
	Other non-functional requirements		Improve the device that monitors laying and compacting temperature data

Table 13. SR Mechanical properties. self-X #4

ID			*AS-SR4.1*
Overall Description			Volumetric/Mechanical properties prediction model
Specific Requirements	Feature	Introduction & Purpose of feature	Based on the real results obtained in the laboratory, generate an intelligent predictive model capable of predicting the volumetric and/or mechanical properties of an untested mixture. As in previous case, not so much historical data are available, so this issue will be manage during the development of the AI model.
		Functional Requirements	Data Exploration: data coming from laboratory Self-X Model Training: prediction of volumetric and mechanical behavior of produced asphalt mixes
	External Interface Requirements	User Interfaces	Intelligent dashboard to support the laboratory operators
		Hardware Interfaces	Not apply
		Software Interfaces	Production data: App Connected Plant Laboratory data: HCLab
		Communications Interfaces	Production data: MQTT Laboratory data: SQL Server database, Mongo DB
	Performance Requirements		Production data: - data reception rate of 800Mb/month - time series & alphanumeric data

			Laboratory data: - data reception rate of 2Mb/year - numeric data
	Other non-functional requirements		

Table 14. SR Gradation characteristic. self-X #4

ID			*AS-SR4.2*
Overall Description			Gradation characteristics verification model
Specific Requirements	Feature	Introduction & Purpose of feature	Verification process of results obtained in the laboratory (for each type of product): Based on the results obtained in the laboratory, and more specifically the gradation of the final product and its bitumen content (real data), a study will be carried out of the possible anomalies arising with respect to the gradation percentages and bitumen content of the asphalt mix design (theoretical data).
		Functional Requirements	Data Exploration: data coming from laboratory Self-X Model Training: algorithm selection and/or model configuration to verify gradation characteristics of asphalt mixes between production and laboratory data.
	External Interface Requirements	User Interfaces	Intelligent dashboard to support the laboratory operators
		Hardware Interfaces	Not apply
		Software Interfaces	Production data: App Connected Plant Laboratory data: HCLab
		Communications Interfaces	Production data: MQTT Laboratory data: SQL Server database, Mongo DB
	Performance Requirements		Production data: - data reception rate of 800Mb/month - time series & alphanumeric data Laboratory data: - data reception rate of 2Mb/year - numeric data
	Other non-functional requirements		

3.2 Steel

Within Steel use case the requirements mapping is as follow:

User stories	self-X Requirements	System Requirement
ST-US1.1	ST-sXR1.01	ST-SR1.1

	ST-sXR1.02 ST-sXR1.03 ST-sXR1.07 ST-sXR1.08 ST-sXR1.09	
ST-US1.2	ST-sXR1.04 ST-sXR1.07 ST-sXR1.08 ST-sXR1.09	
ST-US1.3	ST-sXR1.05 ST-sXR1.07 ST-sXR1.08 ST-sXR1.09	
ST-US1.4	ST-sXR1.06 ST-sXR1.07 ST-sXR1.08 ST-sXR1.09	

3.2.1 self-X #1 solution: "Resilient high-quality raw steel production via EAF"

User stories "Resilient high-quality raw steel production via EAF"	
ST-US1.1	As a scrap yard operator in an steel shop, I want to verify that the scraps are well characterized in terms of chemical composition so that I can load the furnace and keep the final chemistry of the residual elements (like Cu) under control
ST-US1.2	As a scrap yard operator in an steel shop, I want to have a tool for calculating the optimum scrap mix so that I can load the furnace under the overall minimum cost and minimizing the resources
ST-US1.3	As an operator of the furnace, I want to obtain predictions about the temperature and chemistry at the end of the process so that I can anticipate to possible anomalies during the melting process
ST-US1.4	As a data scientist, I want to be able to train predictive model in a simple way so that the time I need to perform a retraining will be minimized

Requirements "Resilient high-quality raw steel production via EAF"	F	NF
ST-sXR1.01. Self-X AI detection must enable the scrap yard operator to identify anomalies by detecting deviations between measured and predicted steel chemistry.	X	
ST-sXR1.02. The self-X- Al diagnose module must allow the scrap yard operator to determine the cause of the anomaly by evaluating the significance of the deviation.	X	
ST-sXR1.03. The Self-X-Al solution must allow the scrap yard operator to update scrap properties using AI-enhanced scrap supervision.	X	

ST-sXR1.04. The Self-X-Al solution must enable the scrap yard operator to Optimize the scrap mix to reduce overall costs through AI-enhanced scrap mix optimization	X	
ST-sXR1.05. The Self-X AI solution must allow the furnace operator to constantly adapt the static AI (and dynamic) furnace model to the new properties of the scrap.	X	
ST-sXR1.06. A self-X AI solution that allows the data scientist to check the prediction accuracy.	X	
ST-sXR1.07. If needed, data scientists, scrapyard, and EAF operators can update the models each heat when the deviation is detected.		X
ST-sXR1.08. The data scientists, scrapyard and EAF operators can optimize the models every 200+ heats.		X
ST-sXR1.09. Operators, engineers, and data scientists must be able to interact with the Self-X AI solution via a web-accessible dashboard and make the final decision.		X

3.2.2 self-X abilities in steel use case prioritization

The proposed holistic solution self-X #1 refers to a toolset for increasing the resilience of raw steel production via EAF. The priority regarding self-X abilities will be focused on detecting unusual situations (anomaly detection) and identifying the root cause of those unusual situations. Moreover, this self-X ability will rely on a repairing mechanism focused on updating and tracking the scrap properties as most critical input variable for the developed modelling tools. These priorities have also been chosen to exploit synergies with the other use cases.

3.2.3 System Requirement

Table 15. SR Steel composition. self-X #1

<table>
<tr><td colspan="3">ID</td><td>ST-SR1.1</td></tr>
<tr><td colspan="3">Overall Description</td><td>The Resilient high-quality raw steel production via EAF is based on the early detection of deviations between measured and pre-dicted/modelled steel composition (CQ) and temperature (T).

The AI based solution will monitor the scrap chemistry so that the Operators, engineers can account with a decision support system, while data scientist receive inputs from an autonomous manner software.</td></tr>
<tr><td rowspan="2">Specific Require ments</td><td rowspan="2">Feature</td><td>Introduction & Purpose of feature</td><td>A holistic approach is necessary in the s-X-AIPI project for the T and CQ resilience obtained by means of the ST-SR1.1;

An evaluation of the deviation significance by means of a diagno-sis module is necessary for the autonomous manner.

Once un update of the scrap properties is calculated, values must be updated so that an optimum mix can be achieved and the fur-nace is loaded under the optimum conditions.

Once the furnace is working a perdition of the final status is needed so that the operator can react to process anomalies.

The autonomous manner will be achieved via anomaly detection; triggers repairing workflows.</td></tr>
<tr><td>Functional Requirements</td><td>• Get historical data
• Validity checks on the inputs
• Perform anomaly detection
• Detect root cause by evaluating the deviation significance</td></tr>
</table>

			• Plan and trigger repairing mechanisms • Save and show results
	External Interface Requirements	User Interfaces	Web interface
		Hardware Interfaces	Server collecting data
		Software Interfaces	Environment for automated workflows
		Communications Interfaces	Data sources and middleware
	Performance Requirements		1. Acquire modelled and measured data 2. Triggered once after every heat
	Other non-functional requirements		…

3.3 Pharma

Within Pharma use case the requirements mapping is as follow:

User stories	self-X Requirements	System Requirement
PH-US1.1	PH-sXR1.01 PH-sXR1.03 PH-sXR1.04 PH-sXR1.05 PH-sXR1.06	PH-SR1.1
PH-US1.2	PH-sXR1.04 PH-sXR1.08	PH-SR1.1
PH-US1.3	PH-sXR1.02 PH-sXR1.07 PH-sXR1.08	PH-SR1.1
PH-US1.4	PH-sXR1.03 PH-sXR1.04	PH-SR1.1
PH-US2.1	PH-sXR2.05	PH-SR2.1
PH-US2.2	PH-sXR2.01 PH-sXR2.02 PH-sXR2.03	PH-SR2.1
PH-US2.3	PH-sXR2.08 PH-sXR2.09	PH-SR2.1
PH-US2.4	PH-sXR2.01 PH-sXR2.02 PH-sXR2.03	PH-SR2.1

	PH-sXR2.04	
PH-US2.5	PH-sXR2.05	PH-SR2.1
PH-US3.1	PH-sXR3.01 PH-sXR3.02 PH-sXR3.03 PH-sXR3.04 PH-sXR3.05	PH-SR3.1

3.3.1 self-X #1 solution: "OCT image quality"

During the process, OCT monitors the suspension and passes these images on to the autonomic manager for further processing. For this, the self-X capabilities "detect" and "diagnose" will be implemented. Self-X detect will perform a check on the image quality and decide whether they can be evaluated in a meaningful way or whether a fault occurred.

User stories "OCT image quality"	
PH-US1.1	As a process operator I want to obtain real-time information about the status of the electrode. The AI solution should present me with the percentage of the surface that is corroded or where particles are sticking (fouling case)
PH-US1.2	As a process operator (not an OCT expert) I want the autonomous manager to inform me if the position of the OCT has shifted and to assist me in readjusting it
PH-US1.3	As a data scientist I want to obtain periodic OCT images of the electrode for further processing
PH-US1.4	As a data scientist I want to obtain the percentages of corrosion or fouling for the development of the process model that can predict the conversion rate trajectory

Requirements "OCT image quality"	**F**	**NF**
PH-sXR1.01. Images can be obtained from OCT 3D sensor	X	
PH-sXR1.02. Images will be transferred to central lab computer via an appropriate protocol (REST-API, gRPC)	X	
PH-sXR1.03. Images can be classified into "good" and "faulty"	X	
PH-sXR1.04. Faulty images can be categorized depending on the fault	X	
PH-sXR1.05. Sample image will be sent to UI for human in the loop	X	
PH-sXR1.06. Image description including fault classification will be sent to UI for human in the loop	X	
PH-sXR1.07. OCT unit must be connected to lab computer		X
PH-sXR1.08. OCT sensor position must be determined correctly		X

3.3.2 self-X #2 solution: "Process ML model prediction"

The biggest component with self-X capabilities is the input data and process model supervision. In the first step, the incoming values from OCT, IR and the power supply are checked for plausibility, both individually and combined. The second part checks the process surrogate model which predicts the expected conversion and concentration as well as expected faults based on data coming from the sensors. Further, drifts in the predictions can be detected and the model can be retrained if needed.

User stories "Process ML model prediction"	
PH-US2.1	As a process operator I would like to see a predicted conversion rate trajectory, which will show me when the process is due to finish (i.e. the conversion is complete)
PH-US2.2	As a process operator I want to be informed in the case of sensor malfunction
PH-US2.3	As an operator I want to monitor the process surrogate model and want to be informed if the predictions do not fit the reality anymore
PH-US2.4	As a process operator I want to be informed if an unexpected scenario occurs, as it might be that I have forgotten something or something needs checking
PH-US2.5	As a data scientist I want to obtain clean data from the sensors. I want to rely on the incoming data for modelling
PH-US2.6	As a data scientist I want to rely in my model's predictions for passing on the right data to the control strategy in the next self-X solution

Requirements "Process ML model predictions"	F	NF
PH-sXR2.01. Deviations in the OCT data can be detected	X	
PH-sXR2.02. Deviations in the IR data can be detected	X	
PH-sXR2.03. Deviations in the data from the power supply can be detected	X	
PH-sXR2.04. Outliers in the combined data can be detected.	X	
PH-sXR2.05. The conversion rate and its trajectory over time can be predicted from the sensor data via a surrogate process model.	X	
PH-sXR2.06. The quality of the surrogate model predictions is checked	X	
PH-sXR2.07. The surrogate model is retrained if required.	X	
PH-sXR2.08. Any value changes will be passed on to the UI for the user in the loop.	X	
PH-sXR2.09. Retraining of model and success or failure will be reported to UI	X	
PH-sXR2.10. All lab equipment (OCT, IR, power supply) need to be connected to the lab computer and be able to communicate with appropriate protocols		X

3.3.3 self-X #3 solution: “Control strategy”

The implementation of self-X optimize targets the control of the power supply. The measured data from OCT, IR and the voltage from the power supply are used to predict the trajectory of the expected conversion rate and concentration. Further, the information about faults that have happened or are predicted to occur soon will be used in the control strategy. With a change of the current, some problems are expected to be able to be prevented or reversed.

User stories “Control strategy”	
PH-US3.1	As an expert operator I want to run my process as optimally as possible, i.e. use as little energy and time as possible without damaging any equipment through for example corrosion

Requirements “Control Strategy”	F	NF
PH-sXR3.01. Is able to set the current in the power supply	X	
PH-sXR3.02. Predicted faults can be addressed and prevented	X	
PH-sXR3.03. Detected faults can be reversed or prevented from getting worse	X	
PH-sXR3.04. Depending on the scenario, the control strategy will set the current automatically or recommend a user action	X	
PH-sXR3.05. Informs the UI for the user in the loop of next steps or recommended actions	X	
PH-sXR3.06. Lab power supply can process input		X

3.3.4 self-X abilities in pharma use case prioritization

Not all self-X components in the pharma use case can be implemented at the same time. Some depend on previous components having been implemented (see Figure 31). In order to address this, the self-X capabilities were divided into three categories: A for critical features, B for non-essential basic features and C for advanced, ‘nice-to-have’ features. Table 16 summarizes this.

Table 16. Prioritization of self-X capabilities.

Component	Phase A: Critical	Phase B: Non-essential Basics	Phase C: Advanced Features
OCT	• Automatic image evaluation • Basic electrode state classification: good/poor	• Refined electrode state classification	• OCT sensor fault classification • Automatically retrain image evaluation if electrode appearance has changed
Process model	• Basic input value check • Prediction of process feasibility: Faults? No conversion?	• Multi-variate input value check • Prediction of measured IR values from process settings, predict process trajectory	• Topological input value check • Automatically retrain model for new chemical reaction or equipment (e.g. electrode material)

Control strategy	• Power off if basic electrode fault occurs	• Adjust power to optimize process • Adjust power to prevent faults	• Adjust power to reverse faults

3.3.5 System Requirement

A description of the self-X managers and the working sequence is given hereinafter. A summary of self-X capabilities and autonomic managers is shown in Figure 31.

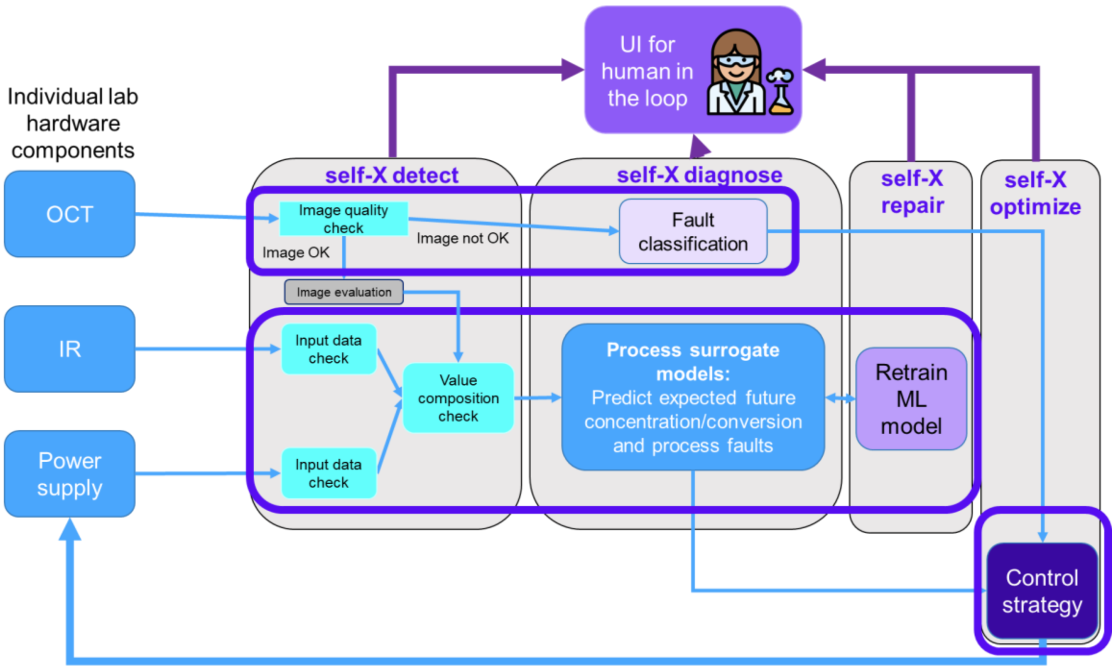


Figure 31. self-X capabilities for electrochemistry. The self-X components are separated in grey boxes, the autonomic managers are inside the purple boxes.

Table 17. SR OCT image quality self-X #1

ID			PH-SR1.1
Overall Description			OCT will be used to monitor the suspension and equipment used in the process. This monitoring will be done in-line during the process with real-time information.
Specific Requirements	Feature	Introduction & Purpose of feature	OCT process monitoring will allow for the detection or prediction of failure as well as the optimization of the process.
		Functional Requirements	OCT images will be generated by the OCT unit. These images will be transferred for image analysis to the self-X component on the lab computer. The self-X detect will decide whether the image quality is sufficient or not. If it is, automatic image evaluation will be implemented to extract numerical values describing the process. If the image is faulty, fault classification will be done.

	External Interface Requirements	User Interfaces	The OCT unit has its own software which will display images. Sample images and values describing the image will be passed on to the UI for the human in the loop
		Hardware Interfaces	The OCT unit will transfer images via REST-API or gRPC if this is sufficiently fast. Alternatively, image processing can be implemented on the OCT unit and the results can be transferred.
		Software Interfaces	Special software for interpreting the OCT sensor data and generating the image will be used on the OCT unit.
		Communications Interfaces	No communication between OCT software and interface for transferring images to the lab computer. A given folder will be checked in intervals for new images and these will be transferred.
	Performance Requirements		SR-RPCE-OCT-1 should provide at least one OCT scan evaluation per minute.
	Other non-functional requirements		

Table 18. SR Process ML model prediction self-X #2

ID			PH-SR2.1
Overall Description			Plausibility check for input and surrogate model for the process linking data from sensors to process performance.
Specific Requirements	Feature	Introduction & Purpose of feature	Expected process performance (conversion rate, expected concentration) can be predicted based on sensor input. This will also provide input for the self-X optimize control strategy.
		Functional Requirements	Input values are checked for all sensors: IR, OCT, power supply. Composition of values is checked for plausibility. The predictions of the surrogate model are compared to the sensor readings. Potential model drifts are detected. The model is retrained if needed (for example, if a different chemical process is run)
	External Interface Requirements	User Interfaces	No direct ones, each component of SR-RCPE-process_model will send information to the UI for the human in the loop.
		Hardware Interfaces	If SR-RPCE-process_model-1 is run on the same computer as the UI for the human in the loop, there are no hardware interfaces. Otherwise, the computers will be connected via Ethernet.

		Software Interfaces	None
		Communications Interfaces	SR-RPCE-process_model-1 will communicate with other autonomic managers via REST-API or gRPC.
	Performance Requirements		
	Other non-functional requirements		

Table 19. SR Control strategy self-X #3

ID			PH-SR3.1
Overall Description			The self-X component for optimization will be implemented via a control strategy. The current in the power supply can be adjusted to allow for a more efficient process or to prevent faults.
Specific Requirements	Feature	Introduction & Purpose of feature	The current during the electrochemical reaction influences process performance. If the current is too low, the reaction progresses very slowly. If the current is too high, equipment failure and other problems might occur. Some problems can be reversed by adjusting the current. Goal of the optimization strategy is to find a setting of the current which allows for a reasonable fast processing time while avoiding common failures. Further, the control strategy will be used to counteract such failures if possible.
		Functional Requirements	Input values from all sensors are available. Current on the power supply can be set automatically. The control strategy will be used to follow a given process trajectory defined by the measured sensor values (IR, voltage, OCT) If a fault occurs or is about to occur, the control strategy will adjust the settings to reverse it or prevent it.
	External Interface Requirements	User Interfaces	No direct ones, each component of SR-RCPE-process_model will send information to the UI for the human in the loop-
		Hardware Interfaces	If SR-RPCE-process_model-1 is run a separate computer than the UI for the human in the loop. The computers will be connected via Ethernet or the internet.
		Software Interfaces	The control strategy will be implemented in Matlab. The component in Matlab will get the data from SR-RPCE-process_model-1 and communicate with the power supply via a REST-API or gRPC.
		Communications Interfaces	SR-RPCE-process_model-1 will communicate with other autonomic managers via a REST-API or gRPC.

	Performance Requirements	**Following a process trajectory:** the quality of this will be determined by the coefficient R^2 that measures the deviation of the measured trajectory from defined ideal process trajectory. **Fault prevention:** the success is determined by the process running with no major faults. Minimal deviations are necessary to detect the potential occurrence of a fault. The control strategy is expected to prevent it. **Fault reversal:** in some cases it might be possible to reverse certain faults by adjusting the settings. The time it takes for the control strategy to do so will be compared with the time it takes a human expert to counteract.
	Other non-functional requirements	

3.4 Aluminium

Within Aluminium use case the requirements mapping is as follow:

User stories	self-X Requirements	System Requirement
AL-US1.1	AL-sXR1.01 AL-sXR1.03 AL-sXR1.05	AL-SR1.1
AL-US1.2	AL-sXR1.01 AL-sXR1.02 AL-sXR1.03 AL-sXR1.04 AL-sXR1.06 AL-sXR1.07	AL-SR1.1
AL-US1.3	AL-sXR1.04 AL-sXR1.05	AL-SR1.1
AL-US1.4	AL-sXR1.02 AL-sXR1.04 AL-sXR1.06 AL-sXR1.07	AL-SR1.1

3.4.1 self-X #1 solution: "Aluminium mix recipe"

User stories "Aluminium mix recipe"	
AL-US1.1	As an aluminium plant operator, I want to know the expected chemical composition from a given recipe so that I can evaluate my preferred combination of materials from a quality perspective

AL-US1.2	As an aluminium plant operator, I want to have up to 3 combinations of scraps for the aluminium recipe based on the total cost, availability of materials, and expected composition of the mixture so that I can select which recipe suits better the heat according to the most influential factors for the enterprise
AL-US1.3	As an aluminium plant operator, I want to have information about potential deviations on the estimated scrap properties based on their behaviour during the process so that I can update those estimations for future recipes
AL-US1.4	As an aluminium general manager, I want to be able to have a proposed planning of recipes according to planned orders so that I can make medium- to long-term process estimations like production scheduling

Requirements "Aluminium mix recipe"	**F**	**NF**
AL-sXR1.01. Aluminium recipes shall result in aluminium products within the acceptable quality requirements specified in the norms	X	
AL-sXR1.02. The IDSS shall provide its solution considering the available materials at the plant and their characteristics		X
AL-sXR1.03. The IDSS shall work in a heat basis approach upon operator's request	X	
AL-sXR1.04. The IDSS shall be able to react to changes in the plant e.g., new raw material available	X	
AL-sXR1.05. The IDSS shall work with uncertainty (missing, incomplete, or unrepresentative information), when possible, exploiting the use of advanced statistics and generative approaches e.g., missing or incomplete information, not representative information from initial chemical analysis of incoming material lots. to ensure the absence of missing values in the processed datasets i.e., zero count of "Not Available - NA" and "Not a Number – NaN" values		X
AL-sXR1.06. The IDSS shall be able to consider and include human feedback, incorporating dedicated human-generated new data in the re-training/optimization processes of the algorithms based on the real recipes selected and specific performance surveys		X
AL-sXR1.07. The solutions proposed by the AI models shall be shown to operators via user-friendly HMIs and personalized notifications according to users' preferences	X	

3.4.2 self-X abilities in aluminium use case prioritization

The IDSS will be the core solution provided in the aluminium use-case (summarized in self-X #1), supported by the development of an aluminium passport for data standardization and exchange. The critical and foremost important aspect of the IDSS will be the support in the general recipe planning, which is currently done in a bi-weekly basis by the general manager based on its expertise. Thus, the IDSS must provide data-driven suggestions to operators to complement and support their decision in the general planning of recipes, exploiting all data available from the different sources in order to enhance human capabilities e.g., considering historical recipes, information from more than 100 silos, previous quality outcomes from melting processes, etc.

The adaptive planning capability of the IDSS will follow and complement the IDSS solution in a second place ("nice-to-have"). One of the key essential points that need support is the alloying process, which, again, it is based on human expertise relying on the chemical analysis performed after primary melting. The IDSS will aim to assist in this decision-making process by considering the recipe used, the chemical analysis results, the norm to be satisfied, and all previous cases provided by historical data. In parallel, the aluminium passport will be developed as a means to provide information about aluminium products throughout their life cycle in a standardized format.

Due to the high dependence on human expertise in current decision-making processes, the incorporation of human feedback is essential in all developments envisaged. Besides that, the IDSS must be able to deal with the uncertainty of the process e.g., using generative approaches to complete and/or enlarge data, as well as be able to self-configure adapting to changes in data distribution e.g., due to human feedback or new materials. To ensure its robustness in real industrial environments, the IDSS must also count with self-protection capabilities to ensure proper behaviour and avoid potential downtimes.

3.4.3 System Requirement

Table 20. SR Aluminium self-X #1

<table>
<tr><td colspan="3">ID</td><td>AL-SR1.1</td></tr>
<tr><td colspan="3">Overall Description

Explain what the software does describes its application, including relevant benefits, objectives, and goals describes general factors affecting it and its requirements</td><td>The IDSS will provide suggestions and recommendations to operators in the plant to support their daily decision-making tasks, focusing on the selection of raw materials to prepare the aluminium recipe. As described in section 2.4.4, the capabilities of the IDSS will be two-folded: on the one hand, it will support the general recipe planning, whereas on the other hand, it will provide adaptive support throughout the process e.g. alloying.
With this solution, the operator will have an objective, data-driven proposition about the optimal recipe based on process data, enhanced by the capability of AI approaches to deal with the intrinsic uncertainty of the process and/or missing data.
In addition, operators are expected to give their feedback to the system to assess the suitability of the proposed recipe, enabling the self-adaptation of the system considering human feedback.</td></tr>
<tr><td rowspan="5">Specific Requirements</td><td rowspan="2">Feature</td><td>Introduction & Purpose of feature</td><td>The main goal of the IDSS is to support operators in the general planning of recipes, as well as provide on-line support during the process.

The IDSS is expected to optimize the scrap collection strategy and overall planning program, meeting all quality standards provided by the norms.</td></tr>
<tr><td>Functional Requirements</td><td>- Validate and integrate available data

- Infer recipes based on the actual scrap stock/or the stock data provided by the user

- Support operators during the process e.g., alloying

- Adapt to changes: new materials available, clients' requests.

- Deal with uncertainty: missing or incomplete data

- Self-optimize over time to adapt to new data and human feedback

- Model version management for self-healing</td></tr>
<tr><td rowspan="3">External Interface Requirements</td><td>User Interfaces</td><td>The main software application will be developed in Python and delivered as a dockerized desktop tool in IDALSA's PC. This application will show the suggested recipes and generate the corresponding reports for the main manager and operators, complementing current practices for recipe planning. The application will request and process human feedback via dedicated forms and surveys through the app.</td></tr>
<tr><td>Hardware Interfaces</td><td>No HW interface required, currently all data is already digitally available.</td></tr>
<tr><td>Software</td><td>- Docker</td></tr>
</table>

		Interfaces	- Python libraries e.g., tensorflow, numpy.
		Communications Interfaces	- No communications interfaces are preliminary required since the main management application at IDALSA already integrates all the information.
	Performance Requirements		The IDSS will have access to the main IDALSA's application that contains the database of all process data, including information about the availability of materials, silo location, historical recipes, quality control, etc. Thus, once trained, the IDSS shall only access the most recent information upon request to perform recipe planning. In a parallel line, the IDSS is expected to self-optimize in time, thus accessing and processing information in the database at specific time slots. For the general recipe planning, the inference of recipes will be done in an order- or periodic- (e.g., biweekly) basis, whereas the adaptive support shall respond in an online basis e.g., once per heat to support alloying.
	Other non-functional requirements		- Portability and compatibility: Compatible with Linux. It will run in external premises. - Reliability and availability: 24/7 - Security: Working under the security restrictions of the client. It will work in client premises. - Usability: User-centered design based on human feedback to be user-friendly.

4 Technical Requirements for the self-X AI technologies

4.1 Asphalt

In this use case there are three existing data infrastructures required for the development of the project: plant and IoT, laboratory, logistics and paving.

Plant production and IoT data: These are obtained through a micro PC box from the SCADA and sent to a WAGO datalogger. This datalogger sends the production data organized in frames through the MQTT protocol to the use case internal data platform. WAGO also has another communication available to be able to send the data frames to another IP address.

Additional IoT sensors (in addition of the ones needed for process control) are directly connected to the WAGO data logger, which also sends the IoT data organized in frames using the MQTT protocol.

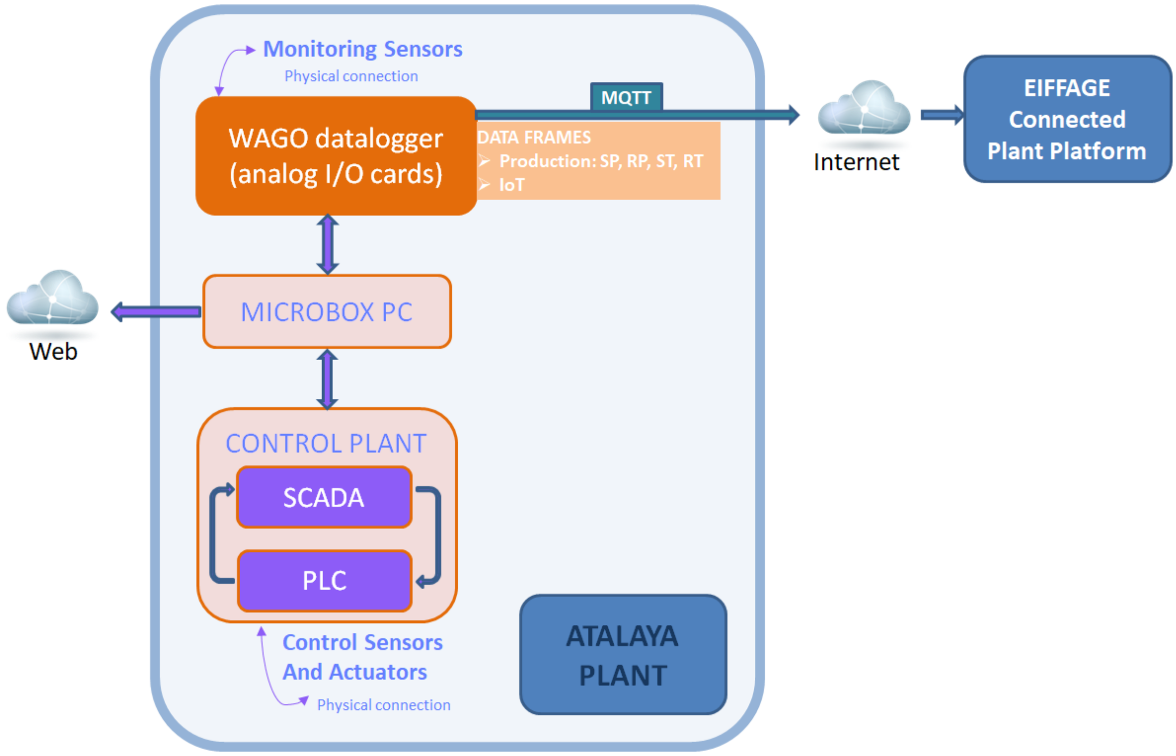


Figure 32. Plant and IoT data infrastructure

Laboratory Data: The lab technician enters the data and analysis results manually into the HCLAB software. HCLAB stores that structured data in a Firebird type database.

These data can be consulted with the HCLAB software, or through SQL queries.

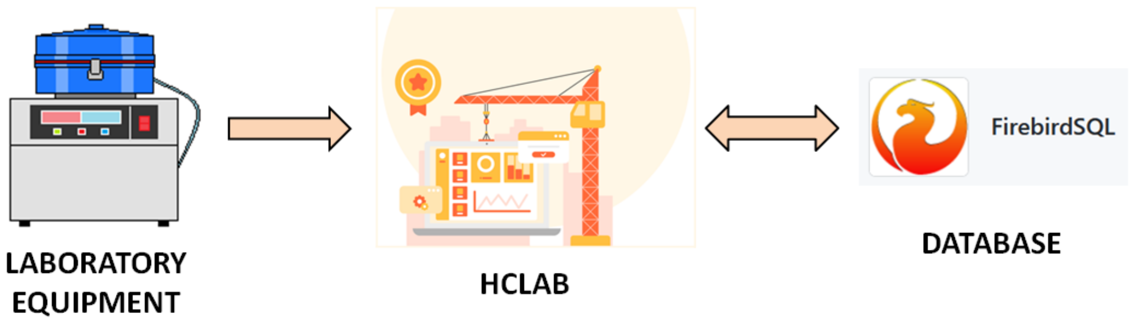


Figure 33. Laboratory data infrastructure

Logistics and paving data: Through existing equipment in the plant, in the trucks, in the pavers and in the compactors, logistics and paving data can be obtained.

Asphalt temperature data is available when loading the truck, as well as when it is paved and compacted. In the pavers and compactors the air temperature and the GPS coordinates of the equipment are also available.

There is also a truck monitored with the temperature of the air and the asphalt, and its GPS coordinates.

In addition, trucks are identified at the plant, both at filling and at the weighbridge, and upon arrival at the paving area, by means of Bluetooth beacons.

All these data are transmitted to a central service (WSIOT) that allows them to be consulted through web browser.

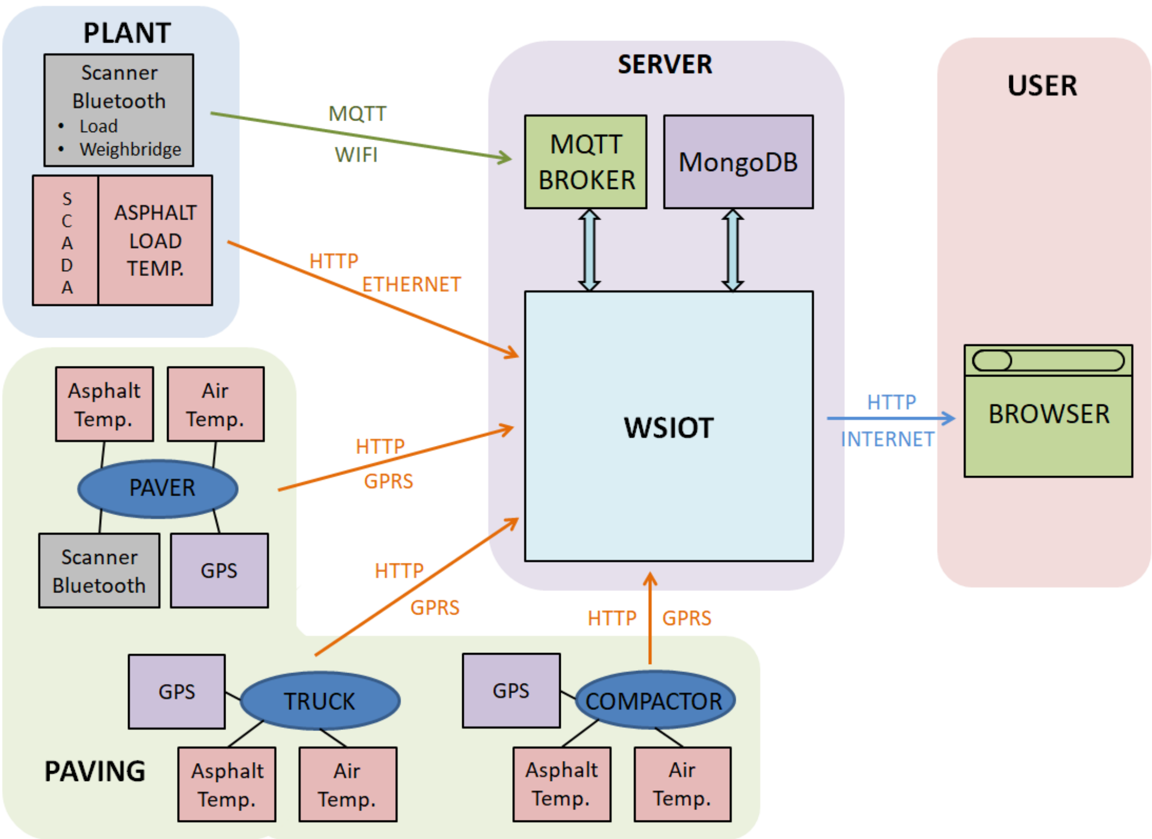


Figure 34. Paving and logistics data infrastructure

4.1.1 Data Sources

Table 21. Asphalt Use Case Data Source #1

Data Involved name			Asphalt Integrated Production Sensors and Actuators
Type			Production Data (both drying and mixing processes)
Details	APIs		WAGO PLC datalogger (MQTT protocol) (from Atalaya EIFFAGE Asphalt plant premises)
	Data	Description	Sensors Data Actuators Data

		Format	REAL, NUMBER
		Data set size	Estimated 300 ~ 600 MB/month (with Ts = 5s) Ts (Sample Period in seconds) (Sample Period, as low as 5 sec for drying process, minutes for mixing process)
		Speed of production	Seconds (drying process) or minutes (mixing process)
		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 22. Asphalt Use Case Data Source #2

Data Involved name			Asphalt Production Setpoints
Type			Production Data (setpoints)
Details	APIs		WAGO PLC datalogger (MQTT protocol) (from Atalaya EIFFAGE Asphalt plant premises)
	Data	Description	Production Data Setpoints
		Format	REAL, NUMBER
		Data set size	Estimated ~2 MB/month (with Ts = 5sec) Ts (Sample Period, as low as 5 sec for drying process, minutes for mixing process)
		Speed of production	Seconds or Minutes
		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 23. Asphalt Use Case Data Source #3

Data Involved name			Production Asphalt mix design
Type			Production Data
Details	APIs		WAGO PLC datalogger (MQTT protocol) (from Atalaya EIFFAGE Asphalt plant premises)
	Data	Description	Production Asphalt Mix Design Data
		Format	STRING
		Data set size	Estimated 2 MB/month (Sample period in minutes for asphalt production process)
		Speed of production	Minutes (usually for both drying and mixing process setpoints)

		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 24. Asphalt Use Case Data Source #4

Data Involved name			Asphalt IoT WAGO PLC Sensors
Type			IoT Sensors Data (Consumed Amps and Energy of the different motors, Different Tanks Levels and Temperature (bitumen, fuel), consumed fuel volume at burner and thermal oil boiler, Motors ON/OFF status, Weather Station data, Humidity at first 2 cold aggregates and RAP bins, …)
Details	APIs		WAGO PLC datalogger (MQTT protocol) (from Atalaya EIFFAGE Asphalt plant premises)
	Data	Description	IoT Sensors Data (physical connection to the WAGO datalogger of different monitoring sensors)
		Format	REAL, NUMBER
		Data set size	Estimated 130~150 MB/month (with Ts = 1sec) Ts (Sample Period, as low as 1 sec/alternating monitoring data)
		Speed of production	Second
		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 25. Asphalt Use Case Data Source #5

Data Involved name			Asphalt Laboratory Analysis
Type			Laboratory Data
Details	APIs		Historical Data: Excel Files Actual Data: Software HCLAB, Firebird database
	Data	Description	Laboratory Analysis of: Aggregates RAP (reclaimed asphalt paving or recycled asphalt from old roads) Asphalt mixes
		Format	REAL, STRING
		Data set size	Estimated 15~30 KB/month
		Speed of production	Hours

		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 26. Asphalt Use Case Data Source #6

Data Involved name			Asphalt Logistics
Type			Truck location data
Details	APIs		Web Service and MongoDB database
	Data	Description	Identifier of the trucks in the plant and in the paving area. Beacon signal power.
		Format	NUMBER, STRING
		Data set size	Estimated 200~400 KB/month (with Ts = 8 min) Ts (Sample Period, as low as 8 min at plant and the paving area)
		Speed of production	Minutes
		Availability to derive an open dataset from it (and under which conditions)	Yes

Table 27. Asphalt Use Case Data Source #7

Data Involved name			Asphalt Paving
Type			Paving Data (temperature and coordinates)
Details	APIs		Web Service and MongoDB database
	Data	Description	Temperature of the asphalt mix in the truck load (at plant) Asphalt mix temperature, air temperature and GPS coordinates in the paving area.
		Format	NUMBER, STRING
		Data set size	Estimated 10~15 MB/month (with Ts = 10 sec) Ts (Sample Period, as low as 10 sec at plant and the paving area)
		Speed of production	Seconds
		Availability to derive an open dataset from it (and under which conditions)	Yes

4.1.2 Available Component

Table 28. Available component #1 Asphalt Use Case: WAGO datalogger

<table>
<tr><td colspan="3">ID</td><td>AC-EIFF-1</td></tr>
<tr><td colspan="3">Responsible partner</td><td>EIFFAGE</td></tr>
<tr><td colspan="3">Tool name</td><td>WAGO PLC datalogger</td></tr>
<tr><td colspan="3">Overall Description</td><td>Each asphalt plant has its own WAGO brand datalogger PLC to send monitoring and production data to an EIFFAGE proprietary cloud system (using MQTT protocol) to remotely visualize and monitor key manufacturing related data used by engineering and management EIFFAGE personnel.</td></tr>
<tr><td rowspan="7">Details</td><td colspan="2">Functionalities offered</td><td>Cloud-based Data Monitoring
Continuous Process Data (cold aggregates drying process)
Event-based (batch asphalt mixing production) Process Data</td></tr>
<tr><td rowspan="3">Data input</td><td>Description</td><td>Production data (sensors, actuators, asphalt mix design, …) from the SCADA control plant system</td></tr>
<tr><td>Format</td><td>UDP frames with proprietary definition (but known) from SCADA plant system provider.</td></tr>
<tr><td>Standard adopted</td><td>Up to 8 or more UTF8 ASCII characters in a UDP frame
• 1st and 2nd character: type of data frame
• 3rd character: data category
• 4th character and 5th character: index number of the data category (01 to 99)
• 6th and 7th character: data number (00 to 99)
• 8th character and following up to the data separator (comma) or end of frame character
For example, RPA1000175 frame means:
• RP: pre-dosing production frame
• A: general parameter
• 1000: data reference = Aggregate dryer outlet temperature
175: data value = 175 ° C</td></tr>
<tr><td rowspan="3">Data output</td><td>Description</td><td>Production data (sensors, actuators, asphalt mix design, …) from the SCADA control plant system
Sensors data from monitoring sensors physically connected to the WAGO datalogger</td></tr>
<tr><td>Format</td><td>Standard JSON payload</td></tr>
<tr><td>Standard adopted</td><td>MQTT protocol</td></tr>
</table>

	Integration requirements	WAGO dataloggers can send data to 2 different IP addresses. One is already in use for proprietary EIFFAGE cloud-based Data Platform. It is needed to setup the 2nd IP address to send the same data to the desired platform/system. Setup at destination system an MQTT broker to subscribe and publish the corresponding data to be integrated in the desired platform/architecture.

Table 29. Available component #2 Asphalt Use Case: HCLAB (LIMS application)

ID			AC-EIFF-2
Responsible partner			EIFFAGE
Tool name			HCLAB (LIMS application)
Overall Description			In the asphalt plant there is a laboratory to obtain the composition, volumetric and mechanical properties of the samples of asphalt mixes, aggregates and RAP. The laboratory analysis are managed through HCLAB software. HCLAB is a Laboratory Information Management System (LIMS). HCLAB stores the laboratory analysis data in Firebird database. Access to this database is possible through either HCLAB or SQL queries.
Details	Functionalities offered		Recording and querying of analysis data using the HCLAB software. Access to analysis data using SQL queries.
	Data input	Description	Data of the analysed samples and of the results obtained from the analyses. These data are manually entered into the HCLAB program by the laboratory technician.
		Format	Alphanumeric data manually entered into the HCLAB software through its data entry interface.
		Standard adopted	The one from the HCLAB software form
	Data output	Description	Data of the analysed samples and of the results obtained from the analyses, manually entered into the HCLAB program.
		Format	Firebird database
		Standard adopted	SQL
	Integration requirements		Develop a small application that performs SQL queries to the HCLAB software's Firebird database and at the same time is an MQTT editor of that data. This application will be located in EIFFAGE installations.

		Setup at destination system an MQTT broker to subscribe and publish the corresponding data to be integrated in the desired platform/architecture.

Table 30. Available component #3 Asphalt Use Case: Truck location

ID			AC-EIFF-3
Responsible partner			EIFFAGE
Tool name			Truck location.
Overall Description			Each truck carries an M2 BLE beacon (Bluetooth) with an assigned name to identify the truck. Using Bluetooth scanners (MOKO Plug MK105) installed in the loading area and on the plant's scale, trucks near these areas are detected and identified. In the paving area, there is a Bluetooth scanner into the asphalt paver (customized ESP32), which allows the truck to be detected an identified in this area. In this way, the loading time of the truck and the arrival time at the paving area are known. Through this system, production data can be related to paving data.
Data:	Functionalities offered		1. Truck loading time. 2. Truck paving area time. 3. Beacon signal power.
	Data input	Description	Dataset: scanner ID, beacon/truck ID, start and end reading time, and Bluetooth signal power. Plant area: The dataset is transmitted via WIFI using the MQTT protocol to the central service (WSIOT). Paving area: The dataset is transmitted via GPRS using the HTTP protocol to the central service (WSIOT). WSIOT stores the dataset in the database (MongoDB) and provides them to the browser when requested.
		Format	Plant area: MQTT messages. Paving area: HTTP messages.
		Standard adopted	Thera are seven values associated with the location data. 1st data: ID scanner (string) 2nd data: ID beacon (string) 3rd data: start time (datetime) 4th data: end time (datetime) 5thdata: mindb (int) 6th: maxdb (int)

			7th: quantums (int)
	Data output	Description	Scanner ID, beacon/truck ID, start and end reading time, and Bluetooth signal power.
		Format	Mongo DB database.
		Standard adopted	BSON (binary JSON)
	Integration requirements		Develop a small application that performs JSON queries to the MongoDB database and at the same time is an MQTT editor of that data. This application will be located in EIFFAGE installations. Setup at destination system an MQTT broker to subscribe and publish the corresponding data to be integrated in the desired platform/architecture.

Table 31. Available component #4 Asphalt Use Case: Asphalt paving

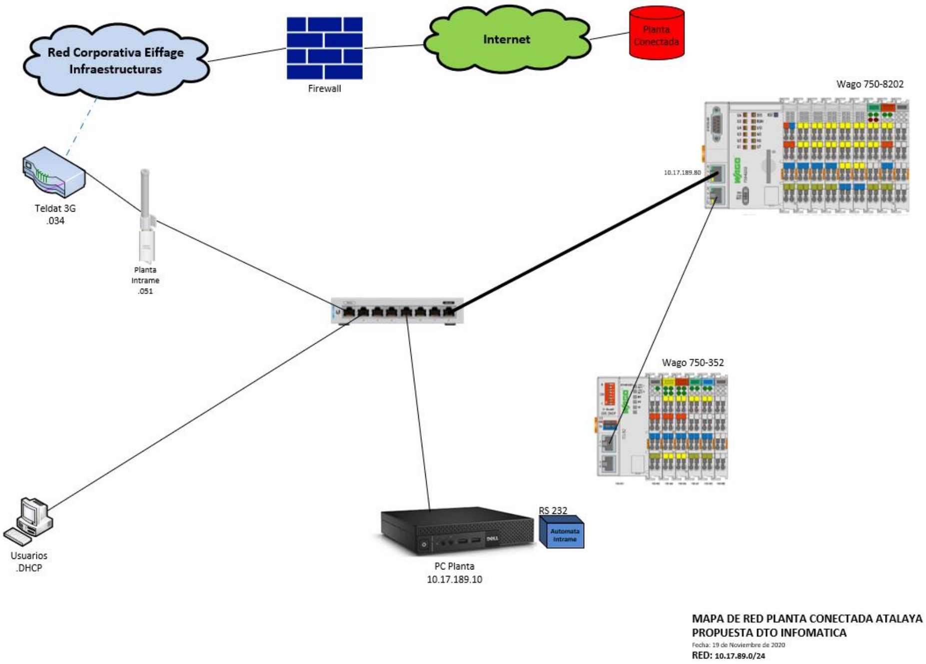
ID			AC-EIFF-4
Responsible partner			EIFFAGE
Tool name			Asphalt paving.
Overall Description			Temperature measurements are available at truck loading (plant), asphalt paver and compactor (paving). Temperature measurements are also available on one of the trucks. Truck loading: the measure of the pyrometer of the mixer is available every 10 seconds. This temperature can be considered as the load temperature. Asphalt paver: air temperature, asphalt temperature (pyrometer), and GPS coordinates are available. Asphalt compactor: air temperature, asphalt temperature (pyrometer), and GPS coordinates are available. Truck: air temperature, asphalt temperature (thermocouple), and GPS coordinates are available, only in one of the trucks.
Data:	Functionalities offered		1. Truck loading temperature. 2. Asphalt paver air and asphalt temperature. GPS coordinates. 3. Asphalt compactor air and asphalt temperature. GPS coordinates. 4. In one truck, air and asphalt temperature. GPS coordinates.
	Data input	Description	Dataset: Temperatures and GPS coordinates. Plant area: the measure of the pyrometer of the mixer is sent by the SCADA using the HTTP protocol to the central service (WSIOT).

			Asphalt paver, asphalt compactor and truck: the measures of temperatures and GPS coordinates are sent by CA/EX/CO devices via GPRS using the HTTP protocol to the central service (WSIOT). WSIOT stores the dataset in the database (MongoDB) and provides them to the browser when requested.
		Format	Plant area: MQTT messages. Paving area: HTTP messages.
		Standard adopted	Plant: Thera are three values associated with the temperature data. 1st data: date of the measure (datetime) 2nd data: ID device (string) 3rd data: asphalt temperature (float) Paving: Thera are nine values associated with the temperature and GPS data. 1st data: date of the measure (datetime) 2nd data: ID device (string) 3rd data: asphalt temperature (float) 4th data: air temperature (float) 5thdata: device battery level (int) 6th: GPS latitude (int) 7th: GPS longitude (int) 8th: GPS altitude (int) 9th: GPS speed (int)
	Data output	Description	Temperatures and GPS coordinates.
		Format	Mongo DB database.
		Standard adopted	BSON (binary JSON)
	Integration requirements		Develop a small application that performs JSON queries to the MongoDB database and at the same time is an MQTT editor of that data. This application will be located in EIFFAGE installations. Setup at destination system an MQTT broker to subscribe and publish the corresponding data to be integrated in the desired platform/architecture.

4.1.3 IT infrastructures

Regarding IT infrastructures available at ASPHALT Use Case, following information has been made directly available by EIFFAGE IT personnel (see Table 32).

Table 32. IT infrastructures, Asphalt Use Case

Topic	Requirement
Deployment	For an on-premises deployment, EIFFAGE currently has a virtualization environment supported by VMWare that could be used if necessary. For an on-cloud deployment EIFFAGE Asphalt plant is open to the use of the Microsoft Azure provider preferably.
Wireless and wired network	There is a local network in the control cabin that is communicated via Wi-Fi antenna with the plant office, which in turn is communicated with the corporate WAN network that has internet access through a firewall. It must be taken into account that the WAN link can be through a connection with a 3G / 4G router, which can limit aspects such as latency, bandwidth and stability of the connection depending on the location of the plant. Although plant history ensures an availability of 99% in most cases.
Security levels	The systems must run within the EIFFAGE corporate network environment and therefore they must meet its standards and rules (the usual of any manufacturing and business network). On the other hand, these standards, rules, and methods could be updated to support these new systems if the change is justified and reasoned and with the corresponding security level according to EIFFAGE security network policies. Neither the controls through trusted IPs, nor those based on trusted MAC addresses are reliable security methods (they can be avoided with very simple techniques today). On the other hand, some of the necessary communications between elements seem to be UDP frames, possibly by broadcast. Therefore, they must be complemented with others based on barriers between networks (firewalls, bridges, etc.), which avoid interference in operations at the most critical point: the manufacturing and control cabin.
Network topology	 **Figure 35: Network map of the asphalt plant** The plant cabin, with all its own devices, is connected to the plant's local network (which includes offices, laboratory, etc.).

	With regard to interconnection with company network and with the Internet, it is practically identical in all plants and is based on a VPN connection (the telephone teldat 3G router acts as a gateway to the outside world [.034] that allows to connect to the corporate servers and other services hosted in the company own cloud or on the Internet. In particular there are several devices related to the sensorization of the plant because it has been used as a prototype in various projects (Smartplant, Connected Plant, etc.) The PC computer running the SCADA manufacturing software is also a PLC PC-based system (SCADA and PLC is an integrated system based on Siemens SIMATIC Box PC) that allows to receive information from the “classic” manufacturing sensors and send orders to the actuators. This PC is also the one that can send the proprietary UDP frames to the local network so that other devices have that information. The list of sensors that are connected to the SCADA system can be obtained from the technical description of the plant installation. Frames may contain information that is not exclusively sensor data, but "more elaborate" data before being sent into the frame (frame format is well documented). Finally, there is a WAGO [.013] datalogger that has been implemented for the EIFFAGE “Connected Plant” project that receives data from another series of sensors and sends them to the central repository of this platform through MQTT protocol.
Network communication	The limitations that may exist, especially with regard to firewalls between sites or to the Internet, could be “relaxed” if the request is justified and no risk is incurred.
Data protection	Any protection related to personal data (taking images or sounds of people), against the leakage of information (data available so that a malicious user can massively take them) or against processes that show signs of assimilating to a generic ransomware or malware (many suspicious file operations).
Data base management	There is no limit to the use of open-source databases. In fact, they are preferred over proprietary solutions. As long as they correspond to recognized products (Postgres, MSSQL, MongoDB, etc.) and maintain a life cycle that includes security updates.
Other	Provide any other information on potential restriction due to your data centre policies that should be taken into account when designing/deploying the self-X AI solution. Some example questions follow: • Any restriction on using Windows, Linux, Mac systems on server side? Preferably Windows for desktop solutions. Optionally Windows or Linux for server applications. • Can existing SW component send information to the s-X-AIPI Platform if deployed outside the factory (over public Internet connection)? It’s possible but it should be documented clearly what information is being sent. • Can the s-X-AIPI Platform send information to a new SW component to be deployed in your premises? Are you using public IP address? It’s possible but it should be clearly documented what information is being sent. It can be enabled a public IP address as long as the port and input protocol are documented, as well as enable a security mechanism. • If you are behind a firewall, which is the available range of ports to be used? The ports will be open on demand once their use is justified and reasoned and it is verified that there are no risks.

	• If you are behind a Reverse Proxy, which are the protocols supported? Are you able to provide a certificate for SSL connections? EIFFAGE IT infrastructure is not behind a ReverseProxy. • Can we use any Open-Source software or do you foresee any license problem? If the software used has a license that is compatible with the project (MIT, GPL, PD, etc.), there is no problem. It should be made a study of the used components and if their license is compatible with the project. • Can we target both Android and iOS devices (if needed) or these devices cannot be connected to your corporate network infrastructure? Currently IT IS NOT ALLOWED the direct connection (via Wi-Fi) of mobile devices to the EIFFAGE corporate internal network.

4.2 Steel

In this use case the s-X-AIPI solution is focused mainly in the scrap supervision and in the prediction of the heat state at end of the electric arc furnace operation. There are two types of available data to deal with:

Acyclic data: Data coming from Manufacturing Execution System (MES). They are defined as process data from a particular heat, where values are not dependent on the time evolution, but fixed values that are stored at different moments of the process.

- Scrap yard related data:
  - Type of all charged materials and amount; scrap availability and price
- Electric arc furnace related data:
  - Final steel weight, temperature and oxygen content; total electrical energy; final chemistry

Cyclic data: Process data coming from PLC collected in real time.

- Electric arc furnace related data:
  - Electrical energy evolution; lime, natural gas and oxygen injections; different type of cooling rates

4.2.1 Data Sources

Table 33. Steel Use Case Data Source #1

Data Involved name			Acyclic data or heat related data
Type			Process data from MES system
Details	APIs		SQL data access
	Data	Description	• Type of all charged materials and amount, t • Steel output, t • Temperature and O2 measurements, ºC / ppm • Total Electrical Energy, kWh • Chemical composition, %wt •…
		Format	Database, integers, floats
		Data set size	Between 100-200 data per heat

		Speed of production	Every hour
		Availability to derive an open dataset from it (and under which conditions)	No

Table 34. Steel Use Case Data Source #2

Data Involved name			Cyclic data or heat related data
Type			Process data from PLC data collector
Details	APIs		SQL data access
	Data	Description	Time series: • Electric energy, kWh • Natural gas and oxygen flow rate, Nm3/h • Injected lime, kg/min • Cooling water input and output, ºC • Cooling water flow rate, Nm3/h • Signals for roof open, slag door open, True/False • …
		Format	Database, integers, floats, booleans
		Data set size	Around 200 data per second
		Speed of production	Every second
		Availability to derive an open dataset from it (and under which conditions)	Maybe (after anonymization)

4.2.2 Available Component

Table 35. Available Component #1 Steel Use Case: User interface

ID			AC-SID-1
Responsible partner			Sidenor
Tool name			s-X AI HMI
Overall Description			Visualization tool for operators with s-X AI technologies. The visualization tool will have different screens dedicated to the users in the loop (Scrap yard Operators, Furnace Operators and Data Scientist)
Details	Functionalities offered		1. User interface
	Data input	Description	• From data sources

			• s-X AI solution
		Format	JSON, Database
		Standard adopted	SQL structured data
	Data output	Description	• s-X AI results
		Format	JSON, SQL data
		Standard adopted	SQL connection
	Integration requirements		Database connections, web server

4.2.3 IT infrastructures

Table 36. IT infrastructures, Steel Use Case

Topic	**Requirement**
Deployment	On premises preferred
Wireless and wired network	Wireless and wired network available
Security levels	VPN, trusted IPs
Network topology	-
Network communication	Communications to outside Company are restricted
Data protection	-
Data base management	-
Other	Can existing SW component send information to the s-X-AIPI Platform if deployed outside the factory (over pubic Internet connection)? **Not recommended** Can the s-X-AIPI Platform send information to a new SW component to be deployed in your premises? Are you using public IP address? **Not possible** Can we target both Android and iOS devices (if needed) or these devices cannot be connected to your corporate network infrastructure? **Not ap-plicable**

4.3 Pharma

In the following subsections the essential data sources and components in the pharma use case are described. Figure 36 summarizes these components and their interaction.

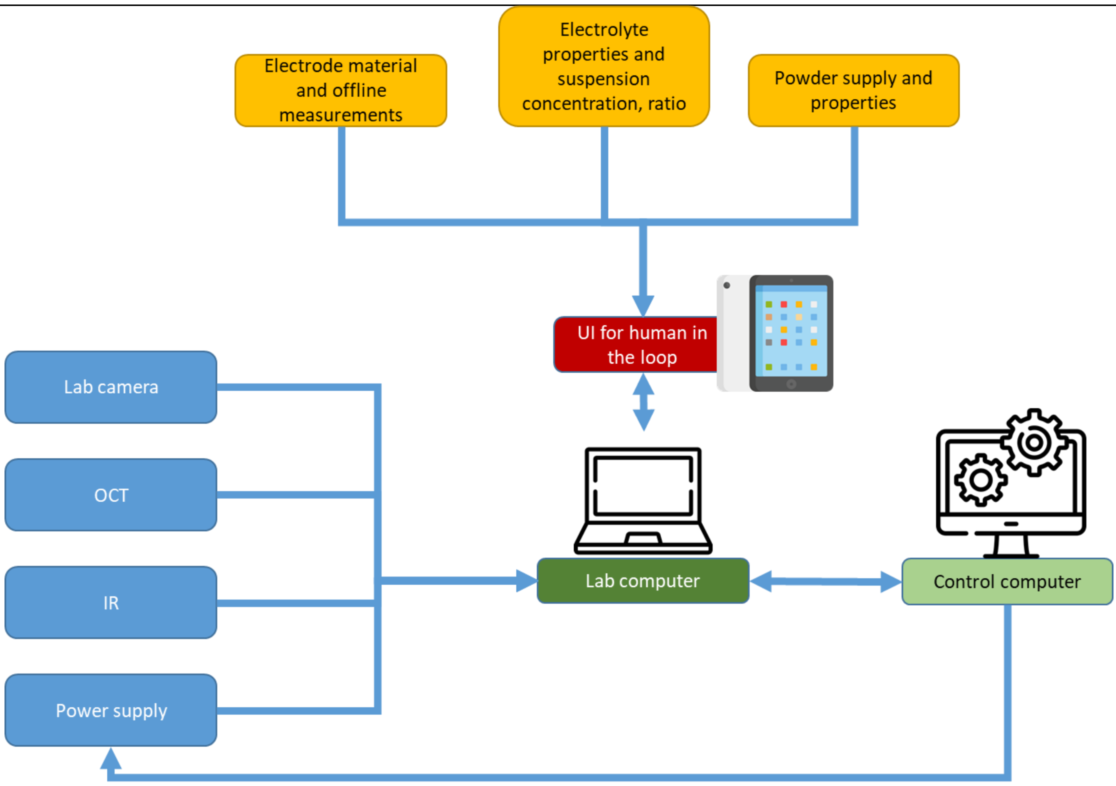


Figure 36. Overview of data sources, components and IT infrastructure in the pharma use case.

- Data sources
  - Supporting electrolyte properties and suspension concentration, ratio
  - Offline measurements of electrode, supplier and material
  - Powder supplier and properties
  - Lab camera: video of cell
  - OCT: 3D images of electrode
  - IR: spectrum of chemical composition
  - Power supply: current and voltage
- Components
  - Lab computer
  - Control computer

4.3.1 Data Sources

The main component of the reaction mixture, apart from the substrate, is the supporting electrolyte. Impurity profiles for the steroid starting material, supporting electrolyte and the solvent are obtained from the specifications data sheet, which may very between lots of the materials. In particular, the presence of impurities that might affect the reaction outcome (e.g., traces of metals) will be highlighted.

This dataset will be manually measured and provided by operator. Most relevant is the particle size and particle size distribution for the subject suspension. This can be readily measured by the operator or a technician by standard techniques such as dynamic light scattering. Variations in the amount of solids added to the mixture due to experimental error will be introduced as solid weight percentage of the suspension.

Table 37. Pharma Use Case Data Source #1

Data Involved name			Electrolyte properties and suspension characteristics
Type			Product data
Details	APIs		None, is manually entered

	Data	Description	• 3D image or 2D slices of the 3D image of the electrode
		Format	Image format
		Data set size	Each 2D slice has around 1 MB, OCT will produce 512 2D slices per minute = 512 MB per minute
		Speed of production	One 3D scan per minute, composed of 512 2D slices
		Availability to derive an open dataset from it (and under which conditions)	A mesh will be calculated out of the 3D images to obtain a representation of the state of the electrode. This will be 512x1024 double values. These meshes will be shared together with example images.

The electrode dimensions and weight will be checked by the operator before the process. The information about the supplier and material will also be provided.

This data is manually obtained and needs to be entered into the UI by the human in the loop before the process.

Table 38. Pharma Use Case Data Source #2

Data Involved name			Electrode properties. Offline measurements of electrode, supplier and material
Type			Product data
Details	APIs		None, is manually entered
	Data	Description	Metadata describing the electrode
		Format	Combination of strings and numeric values
		Data set size	< 1000 bytes
		Speed of production	Is entered manually by the operator before the process starts
		Availability to derive an open dataset from it (and under which conditions)	Very likely

The data will be entered manually by the operator.

Table 39. Pharma Use Case Data Source #3

Data Involved name			Powder supplier and properties
Type			Product data
Details	APIs		None, will be manually entered into the UI
	Data	Description	Metadata describing the electrode
		Format	Combination of strings and numeric values
		Data set size	< 1000 bytes
		Speed of production	Is entered manually by the operator before the process starts

		Availability to derive an open dataset from it (and under which conditions)	Very likely

The lab camera will be pointed at the process and monitor primarily the fill level of the electrolysis cell. Additionally, other features can be extracted from the video.

This data will be obtained via a direct connection to the lab computer.

Table 40. Pharma Use Case Data Source #4

Data Involved name			Lab Camera
Type			Process monitoring data
Details	APIs		USB, vendor provided format
	Data	Description	Video of electrolysis cell
		Format	Video stream as image sequence
		Data set size	To be determined
		Speed of production	To be determined
		Availability to derive an open dataset from it (and under which conditions)	Very likely fully open, videos or snapshots of images can be provided.

The OCT is a light based system using a long wavelength and can hence penetrate certain materials. In this use case, it will be applied to monitor the electrode in the suspension during the process. Due to the nature of OCT, this can also occur through opaque suspensions with particles.

This dataset will be obtained through in-line monitoring. A 3D OCT sensor is constructed and the 3D images are generated by the OCT unit. If performance allows, the 3D images will be transferred to the lab computer and processed on it. If the transfer speed is too slow, the image processing will be done on the OCT unit and the resulting values will be transferred.

Table 41. Pharma Use Case Data Source #5

Data Involved name			OCT (Optical Coherence Tomography) images.
Type			Process monitoring data
Details	APIs		gRPC
	Data	Description	3D image or 2D slices of the 3D image of the electrode
		Format	Image format
		Data set size	Each 2D slice has around 1 MB, OCT will produce 512 2D slices per minute = 512 MB per minute
		Speed of production	One 3D scan per minute, composed of 512 2D slices
		Availability to derive an open dataset from it	A mesh will be calculated out of the 3D images to obtain a representation of the state of the electrode. This will be

		(and under which conditions)	512x1024 double values. These meshes will be shared together with example images.

Infrared (IR) spectroscopy is a standard analytical technique that measures the interaction of infrared radiation with a material, which can be a solid, liquid or gas. The technique is based on the vibrational energies of different functional groups present in the material. Thus, depending on the frequency at which the absorption is measured, the technique can measure specific components in a mixture. In the case of a chemical reaction, as chemical bonds and formed and/or cleaved, IR absorption at the vibrational energy corresponding to those bonds is monitored.

IR probes can be directly immersed in a reaction solution. Alternatively, the solution can be continuous pumped through flow-through monitoring cells.

The IR sensor is connected to the lab computer via USB. As communication protocol, OPC-UA is used.

Table 42. Pharma Use Case Data Source #6

Data Involved name			IR sensor
Type			Process data
Details	APIs		USB, vendor provided format
	Data	Description	Absorbance at a specific wavelength
		Format	Double numbers
		Data set size	Full spectra ca. 1 Mb; monitoring of a particular wavelength few Kb
		Speed of production	Data can be obtained every 15 seconds
		Availability to derive an open dataset from it (and under which conditions)	Yes

The power supply sets the current for the electrochemical reaction. It can also read the voltage. This is a crucial part for having a reaction and making the process work.

Access is via an RS232 to USB connector. The power supply can be addressed with RS232 commands.

Table 43. Pharma Use Case Data Source #7

Data Involved name			Power supply
Type			Process data
Details	APIs		RS232 interface
	Data	Description	• current • voltage
		Format	Double numbers
		Data set size	Two double numbers upon request
		Speed of production	Can be read upon request, we are aiming at one reading per second (two double numbers for current and voltage per second)
		Availability to derive an open	Yes

		dataset from it (and under which conditions)	

4.3.2 Available Components

This is the central unit collection all data and providing information for the UI for the human in the loop. On this unit, most self-X components will be implemented and data will be stored centrally, or passed on to a central data storage unit.

Table 44. Available Component #1 Pharma Use Case: Lab Computer

ID			AC-<RCPE>-<1>
Responsible partner			RCPE
Tool name			Lab computer
Overall Description			Laptop collecting all data sources and components and executing most self-X components
Details	Functionalities offered		1. Data analysis 2. Data for user interface or provides the user interface directly 3. Physical connection to data sources
	Data input	Description	• Sensor data • Static data from the UI provided by the human in the loop • Feedback from the control computer
		Format	Images, double arrays, …
		Standard adopted	OPC-UA, gRPC, RS232
	Data output	Description	Double arrays or processed images, 3D meshes for electrode surface
		Format	Images, double arrays
		Standard adopted	UPC-UA, gRPC, RS232
	Integration requirements		Connection to network, internet

This computer will have the self-X optimize strategy implemented which includes a control strategy to adjust the power supply according the process monitoring information. It will need a Matlab software on it including the control strategy.

Table 45. Available Component #2 Pharma Use Case: Control Computer

ID	AC-<RCPE>-<2>
Responsible partner	RCPE
Tool name	Control Computer
Overall Description	Computer communicating with the lab computer and the power supply to adjust settings and optimize the process

Details	Functionalities offered		1. Communication with lab computer and power supply 2. Controls the power supply to optimize the process
	Data input	Description	• Information from the lab computer about process
		Format	Double arrays
		Standard adopted	gRPC
	Data output	Description	Double arrays or processed images, 3D meshes for electrode surface
		Format	Images, double arrays
		Standard adopted	UPC-UA, gRPC, RS232
	Integration requirements		Connection to internet, local network (VPN)

The user interface for the human in the loop will allow for the display of information as well as processing user input and hence facilitate interaction. This user interface could either be displayed on the lab computer or via a web server run on the lab computer on any other device in the network, such as a tablet for the operator for example.

Table 46. Available Component #3 Pharma Use Case: User interface for human in the loop

ID			AC-<RCPE>-<3>
Responsible partner			RCPE
Tool name			User Interface
Overall Description			Displays process information and processes user input where required
Details	Functionalities offered		1. Display of information about the process 2. Tracking of the process 3. User can enter information prior to process 4. User can interact with system where needed
	Data input	Description	• User enters data before the process begins (such as material data, electrode measurements…) • When the self-X systems needs input the user can react
		Format	Data prior to process in the form of strings for supplier information and numeric for measurements Reaction to certain events, yes/no or input of values
		Standard adopted	gRPC
	Data output	Description	Displays images, numbers, plots
		Format	Images, double arrays
		Standard adopted	Graphical user interface
	Integration requirements		Connection to WIFI

4.3.3 IT infrastructures

Table 47. IT infrastructures, Pharma Use Case (1/2)

Topic	Requirement
Deployment	Lab laptop with main part of self-X solutions located at chemistry lab inside the university of Graz
Wireless and wired network	Ethernet will be used between computers
Security levels	Trusted MAC addresses
Network topology	Part of the university network. All required computers will be added to the university network and availability of access between computers is regulated via the university firewall.
Network communication	gRPC
Data protection	Access to computer controlled via TeamViewer with password
Data base management	No restriction for research purposes. We will choose an open source database.
Other	Both computers, the lab computer and the computer for the control strategy operate on Windows. On the lab computer we could potentially switch, but the control strategy needs Matlab for which Windows is the preferred choice. For self-X components ports 8000-8999 will be used for gRPC. Any restriction on using Windows, Linux, Mac systems on server side? No restrictions, per default no admin privileges for regular users Can the s-X-AIPI Platform send information to a new SW component to be deployed in your premises? Are you using public IP address? Range of TU Graz 129.27.217.x can be accessed from the outside. Ports can be opened per demand. However, computers joining externally into the TU Graz network via VPN are not accessible via this. A rerouting via a different computer can be used for this. If you are behind a firewall, which is the available range of ports to be used? The ports will be open on demand once their use is justified and reasoned and it is verified that there are no risks. RCPE sits behind a Fortigate Firewall. Certain ports can be opened on demand if needed. If you are behind a Reverse Proxy, which are the protocols supported? Are you able to provide a certificate for SSL connections? RCPE uses self-signed Certs in-house. Certified certificates could be bought if required. Can we use any Open Source software or do you foresee any license problem? No restrictions at RCPE.

Table 48. IT infrastructures, Pharma Use Case (2/2)

Topic	Requirement
Deployment	Computer for control strategy located at RCPE main site inside the network of the Technical University of Graz (different form the university of Graz).
Wireless and wired network	Ethernet will be used between computers.

Security levels	Trusted MAC addresses
Network topology	Part of the university network. All required computers will be added to the university network and availability of access between computers is regulated via the university firewall.
Network communication	gRPC
Data protection	None necessary
Data base management	No restriction for research purposes. We will choose an open source database.
Other	Both computers, the lab computer and the computer for the control strategy operate on Windows. On the lab computer we could potentially switch, but the control strategy needs Matlab for which Windows is the preferred choice. For self-X components ports 8000-8999 will be used for gRPC. Any restriction on using Windows, Linux, Mac systems on server side? No restrictions, per default no admin privileges for regular users Can the s-X-AIPI Platform send information to a new SW component to be deployed in your premises? Are you using public IP address? Range of TU Graz 129.27.217.x can be accessed from the outside. Ports can be opened per demand. However, computers joining externally into the TU Graz network via VPN are not accessible via this. A rerouting via a different computer can be used for this. If you are behind a firewall, which is the available range of ports to be used? The ports will be open on demand once their use is justified and reasoned and it is verified that there are no risks. RCPE sits behind a Fortigate Firewall. Certain ports can be opened on demand if needed. If you are behind a Reverse Proxy, which are the protocols supported? Are you able to provide a certificate for SSL connections? RCPE uses Selfsigned Certs in-house. Certified certificates could be bought if required. Can we use any Open Source software or do you foresee any license problem? No restrictions at RCPE.

4.4 Aluminium

The current data acquisition resources for process control at IDALSA are partially available as 2 systems considering the two main burners, namely TRF-1 and TRF-2. Both burners have an analog control panel through buttons with IO that control the operations of feeding, inclination, fusion, etc., with Allen-Bradley 550 panels. Besides that, the burners also have:

- TRF1: Analog control panel with pushbuttons (power supply, burner, oven rotation inclination). PLC model ALLEN BRADLEY SLC 5/04, reports continuous graphic data, in winncc32 through a screen, of hydraulic pressure, instantaneous gas consumption and filter temperature. It also records these data and others on the state of the oven in a .DBF file. The current SCADA only partially registers and presents data but does not control.
- TRF2: Here, in addition to the control panel with push buttons, ALLEN BRADLEY SLC 550 and HMI SIEMENS, that shows the gearmotor amperage curve, dust collection system without data recording.

Regarding the quality and monitoring of the alloy, there is a mass spectrometer for solid metals that records the composition in percentage by weight of the aluminium test tubes during production. Furthermore, the movements of materials within the plant are carried out with mechanical shovels that transport the materials from their storage areas to the furnace's chargers, counting with a weight balance to measure the materials. Figure 37 shows the main components of the control and monitoring architecture in IDALSA.

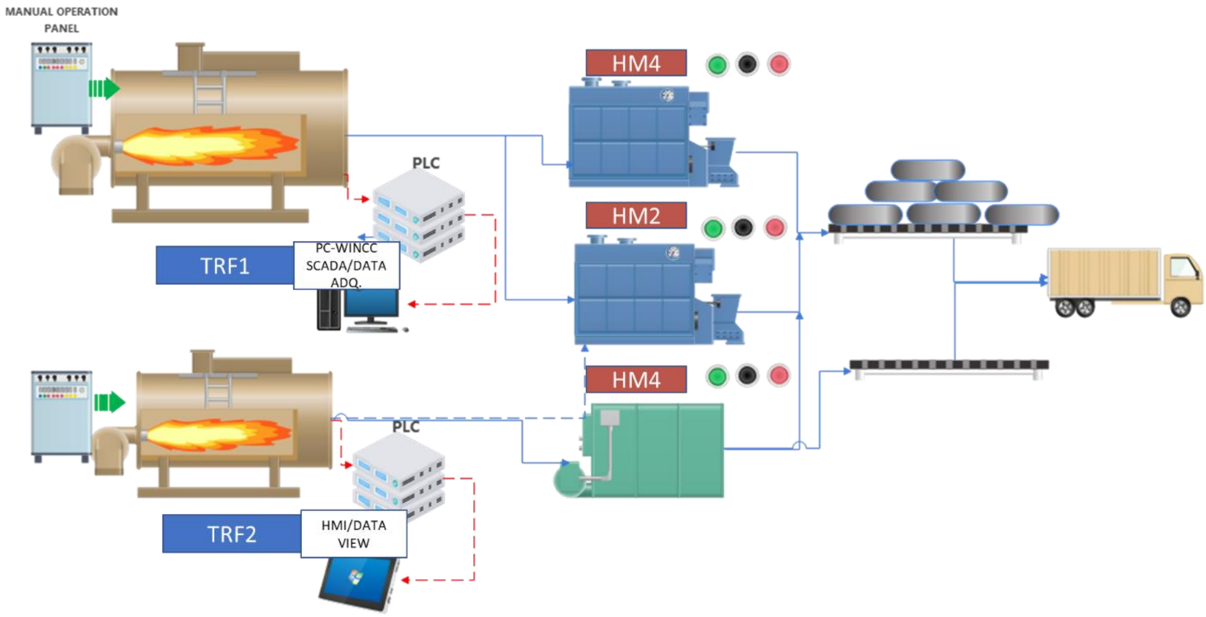


Figure 37. Control and monitoring architecture in IDALSA

IDALSA currently has a production, purchasing and logistics management software based on a LINUX environment connecting the different areas through an Ethernet network (Figure 37). In this software, the production data, purchase of raw materials, quality, and yields are recorded from two main points: finished materials and logistics control. The loading of operational control, production, quality and logistics data is carried out mainly from the fusion booth, where the operators record the processes, measurements and determined quantities and the rest of the inputs are carried out from the administration and purchasing office.

To access the data, the user identification is initially carried out and the platform is navigated by accessing different menus using predefined keys and codes. Here you can enter the data in turns for each heat, material consumed, quantity and type of alloy produced, main energy consumption, additives and also generate predefined reports with current data, allowing traceability of production, quality, stocks and materials produced.

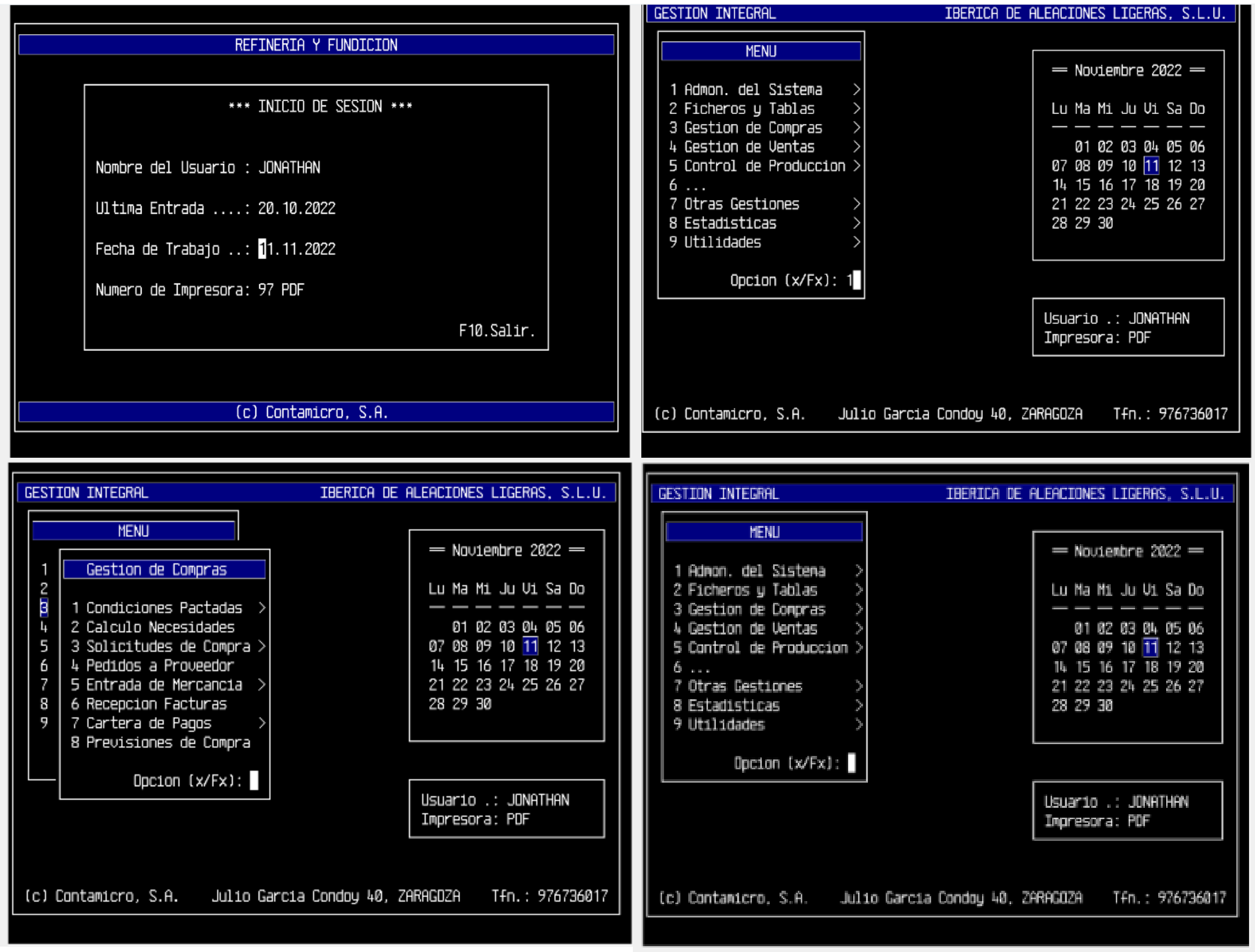


Figure 38. Screenshots of the LINUX application at IDALSA

The most relevant information of the process is integrated in the main LINUX application, and, thus, the IDSS will mainly rely on this data. In order to provide adaptive process support, access to intermediate quality controls during alloying might be required, which might not always be integrated in the main LINUX application, hence, access to the spectrometer data is required. The spectrometer is described in Available Component.

4.4.1 Data Sources

Table 49. Aluminium use case data source #1

Data Involved name:			Management software in Linux
Type: **Software**			The management software was developed in Linux by the company Contamicro. The application keeps record of all the relevant information of the aluminium process in IDALSA at the different stages of the production.
Details	APIs		*SQL and csv data* .
	Data	Description	This software stores all the relevant information of the production at IDALSA, including purchases, sales, material stock, process information (recipes), quality analysis, and gas consumption.
		Format	TXT, PDF, CSV.
		Data set size	100 MB to max size for reports

		Speed of production	The information is updated and validated by the manager in the application in a daily basis. The information available is discrete (available at a heat-basis, order-basis, reception-basis, etc.) and covers all the relevant information of the production process.
		Availability to derive an open dataset from it (and under which conditions)	Yes, but limited – currently there are different user rights to access the database and the datasets are created separately according to their nature e.g., reception phase, melting phase, client requirements, etc.

4.4.2 Available Component

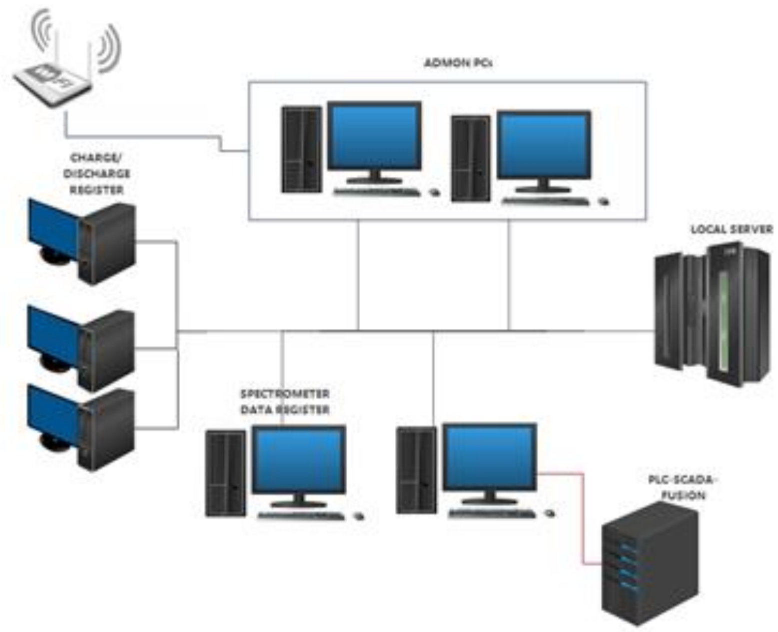
Table 50. Available Component Aluminium Use Case #1: ARL Optical Emission Spectrometer Model MA 276

ID			ARL Optical Emission Spectrometer
Responsible partner			IDALSA
Tool name			ARL Optical Emission Spectrometer Model MA 276
Overall Description			High performance OES spectrometry platform based on photomultiplier tube (PMT) and CCD optics. With this equipment, the chemical composition of an aluminum test tube obtained by fusion of a sample of the input material is measured in order to know how to define the mixtures according to the standard to be met.
Details	Functionalities offered		1. Chemical composition of aluminium alloys 2. WinOE 3. Data analytics reports visualization
	Data input	Description	Sample of the melted aluminum mixture obtained by the fusion of the input materials introduced as a test tube. Test tube with Vs=60 cm3
		Format	Test tube
		Standard adopted	ASTM E1251
	Data output	Description	Provides the weight percentages of metals obtained such as: Fe, Si, Mg, Mn, Cu, Cr , Pb, Ni, Sn, Ti, Zn, Sb, Ca, Na, B, P, V, Zr, Sr, Li.
		Format	DBF, CSV, TXT
		Standard adopted	Pure test tube metallic aluminum 99%
	Integration requirements		Currently, it is not integrated in the Linux management software, and it is available as a stand-alone application in the laboratory at the main industrial facility. Only the final chemical analysis is available at the LINUX application.

4.4.3 IT infraestructures

Table 51. IT infrastructures, Aluminium Use Case

Topic	Requirement
Deployment	Preferably, s-x-AIPI solutions should be deployed on premises for confidentiality issues in order to avoid additional security considerations to transfer information to external

	premises, as well as to be compliant with current requirements and specifications of the SW/HW at the plant. IDALSA currently has a local network and server, but it does not have a cloud service of any kind (open or private access). Authorization for the installation of a system hosted on the Internet can be requested from the address.
Wireless and wired network	The communication at IDALSA is at home, in some points of the ship (scale, spectrum, office) by Ethernet, and administration via ethernet/Wireless (the oven control cabin does not have an internet network, the PC of the TRF1 is only inside the local network/Ethernet). The fusion operator has limited access to the LINUX application to introduce process data and perform minor validation checks of material stocks. The app can make different reports and save them in an internal server.
Security levels	The security is through MAC address
Network topology	The figure below depicts the network topology in IDALSA 
Network communication	PC-SCADA TRF1 with outdated operating system (winxp) as well as motherboard, processor, ram etc., so installing new applications that need more powerful resources will surely be a problem, similar to pc spectro, TRF2 does not have acquisition of data/PC, equally alloyed (without data acquisition or control). Similar situation at different points of the plant and the fusion operator can have accesses at LINUX app by Ethernet red with restrictions to obtain information. The envisaged solution will initially be deployed at the local PC where the LINUX app is located with granted access to the fusion operator to support recipe planning, notifications and alarms might be given via the corporate e-mail or HMIs.
Data protection	It is critical to ensure the anonymization of clients, orders, raw materials, recipes, and production planning to avoid compromising IDALSA's expertise and practices.
Data base management	The database is currently being updated and the conditions and proceedings for its access are still under discussion.
Other	There are not strict restrictions regarding the operating system in which the solution must be developed, but LINUX is preferred for consistency with the current application or WINDOWS due to its general availability at the local PCs in IDALSA.

5 Requirements for Self-X AI adoption in Process Industry

s-X-AIPI project looks to apply a new approach to build AI based applications for the process industry. The approach looks for more autonomous AI applications from the AI creating process itself. For that purpose, in this section requirements of the industrial use cases are matched with overall methodology.

The project methodology intends to apply an AI data pipeline suitable for AI online operation and autonomic behaviour:

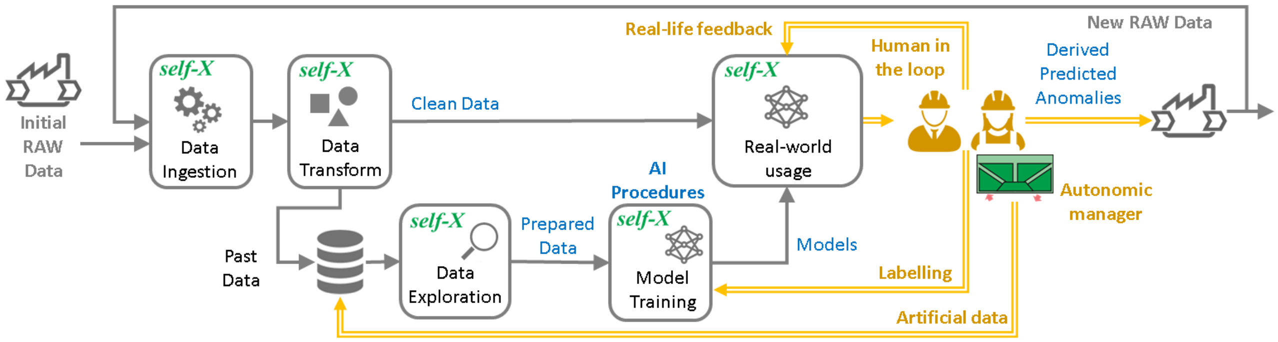


Figure 39. AI data pipeline

Autonomic behavior is based on the application of the autonomic computing paradigm to AI creation process where the different pipeline components exhibit certain (self-X) abilities like self-Healing or self-Configuration. In addition to this s-X abilities, it is proposed to include what is called an autonomic manager that coordinates the different pipeline components.

In order to a clarification for the requirements provided in use cases (self-X detect/diagnose/…), details about the characteristics of an autonomous system from the literature are given hereinafter.

Computing systems that incorporate algorithms that replicate **human-like intelligent capabilities such as learning and self-adaptation**, allow the machine to autonomously operate in certain complex and continuously changing environments, or processes that adopt computational models of intelligent behavior to solve complex problems that humans are not able to solve[4].

AI-based autonomy specifically refers to the ability of an AI-based autonomous system to perform specific tasks independently. They can exhibit unique machine behaviours and evolve to gain certain levels of human-like cognitive, self-executing, and self-adaptive abilities. As AI technology advances, these autonomous systems can be developed to have these capabilities over time. They may successfully operate under some situations that are possibly not fully anticipated, and the results may not be deterministic[5]. Current AI technology only allows to develop autonomous systems that typically perform limited tasks in specific situations. There will be situations that the designers had never considered and which the system cannot handle. For example, no current autonomous vehicles can "solve" all driving scenarios without human intervention.

Although there are many characteristics or properties of autonomous systems, these characteristics are the most important: *self-configuring, self-healing, self-optimizing and self-protecting*[6].

The major goal of autonomous systems is self-management which is only possible when all the components of the system work together to achieve it.

Autonomic manager manages the managed element which can be software or hardware. It uses MAPE (Monitor, Analyse, Plan and Execute) function to do so. The monitor function monitors and collects details from the managed element, analyse function analyses the collected details and check if any change is to be

[4] Wienrich, C., & Latoschik, M. E. (2021). eXtended Artificial Intelligence: New Prospects of Human-AI Interaction Research. arXiv preprint arXiv:2103.15004.

[5] Kaber, D. B. (2018). A conceptual framework of autonomous and automated agents. Theoretical Issues in Ergonomics Science, 19, 406-430

[6] M. Parashar and S. Hariri, "Autonomic computing: An Overview", Springer Verlag, pp. 247-259, 2005.

made in the managed element or resource then plan function develops the necessary actions that are required to be taken these actions are performed using execute function. Knowledge Source can be implemented using database, dictionary or repository. It consists of knowledge that autonomic manager uses to perform self-management tasks, knowledge can be either passed to autonomic manager through policies or is retrieved from external knowledge source or is created itself by autonomic manager[7]. Changes are made in the managed element using policy, which is provided to the autonomic manager. Autonomic manager communicates using touch points, which consist of sensors and effectors. Sensors sense the information about the current state of the managed element and effectors checks the changes that are made to manage element by autonomic manager.

The most important interpretations of the characteristics of autonomous systems are given as follows:

Self-Configuration

- A process by which computer systems automatically adapt their own configuration of components without human direct intervention.
- An autonomous computing system must be able to install and set up software automatically. To do so, it will utilize dynamic software configuration techniques, which means applying technical and administrative direction and surveillance to identify and document the functional and physical characteristics of a configurable item. Also to control changes to those characteristics, to record and report change processing and implementation status, and to verify compliance with specified service levels.
- By this property autonomic system is able to add new components, configure its components, remove old or faulty components, and reconfigure them by themselves with a little or no human interference. These components will be able to adapt the IT environment and would provide users with desired performance and quality

Self-Healing

- Having the power or property of healing one’s self or itself. Autonomic computing refers to the ability of systems to self-diagnose and self-heal without the need for operator intervention.
- Self-healing is a technique that aims to detect, analyze, and repair existing faults within the system.
- Moreover, self-healing system should have the ability to modify its own behavior in response to changes within the environment.
- Self-Healing Autonomic system can heal itself and its components through this property. It is able to detect faulty components, try to diagnose these components using some corrective mechanism. But self-healing process should not harm other components of the system. ROC (Recovery Oriented Computing) is the application for performing self-healing process; it provides us with various mechanisms for recovery of the system from many failures.

Self-optimization

- The ability of the IT infrastructure to efficiently maximize resource allocation and utilization to provide service for both system users and their customers. In the short term, self-optimization primarily addresses the complexity of managing system performance.
- In the long term, self-optimizing software applications may learn from experience and proactively tune themselves in an overall business objective context.
- Workload management uses self-optimizing technology to help optimize hardware and software use and verify that service-level goals are being met. Predictive analysis tools provide views into performance trends, allowing proactive action to be taken to help optimize the IT infrastructure before critical.

[7] IBM white paper, “An architecture blueprint for autonomic computing”, page 1-32, 2003.

- It means autonomic system can optimize itself and its components for improving its efficiency, tune its resources and satisfy requirements of different users. Resource utilization and workload management are aspects of these characteristics.

Self-Protecting

- A self-protecting IT environment must take appropriate actions automatically to make itself less vulnerable to attacks on its runtime infrastructure and business data. These attacks—which often occur on a daily basis—can take the form of unauthorized access and use, malicious viruses that can reformat hard drives and destroy business data, and denial-of-service attacks that can cripple critical business applications. A combination of security management tools and storage management tools are necessary to deal with these threats. Security management tools can help businesses consistently enforce security and privacy policies, help reduce overall security administration costs, and help increase employee productivity and customer ...
- As biological systems are capable of protecting themselves from dangers and other natural calamities similarly autonomic systems have self-protecting property, by this property they are capable of defending themselves and their components from various malicious attacks.
- Autonomic systems can detect hostile behaviors; take various corrective actions for guarding themselves from any attack.

In the following table, the mapping between the Autonomic computing characteristics and the Quality factors[8], which are important from the implementation point of view is given:

Quality Factors	Autonomic computing characteristics	
	Major	Minor
Functionality	Self Configuring Self Optimizing Self Protecting	Self Aware Context Aware
Reliability	Self Healing Self Protecting	---
Usability	Self Configuring	---
Efficiency	Self Optimizing	Anticipatory
Maintainability	Self Configuring Self Healing Self Optimizing	Anticipatory
Portability	Self Configuring	Open

[8] Singh et al, SURVEY ON CHARACTERISTICS OF AUTONOMOUS SYSTEM, International Journal of Computer Science & Information Technology (IJCSIT) Vol 8, No 2, April 2016

These quality characteristics define the requirements for the realization of the use cases functionalities: SELF-X detect, SELF-X diagnose, SELF-X repair and self-X Optimization.

Following figure illustrates the causes/triggers for activating autonomous behaviour of the AI pipeline.

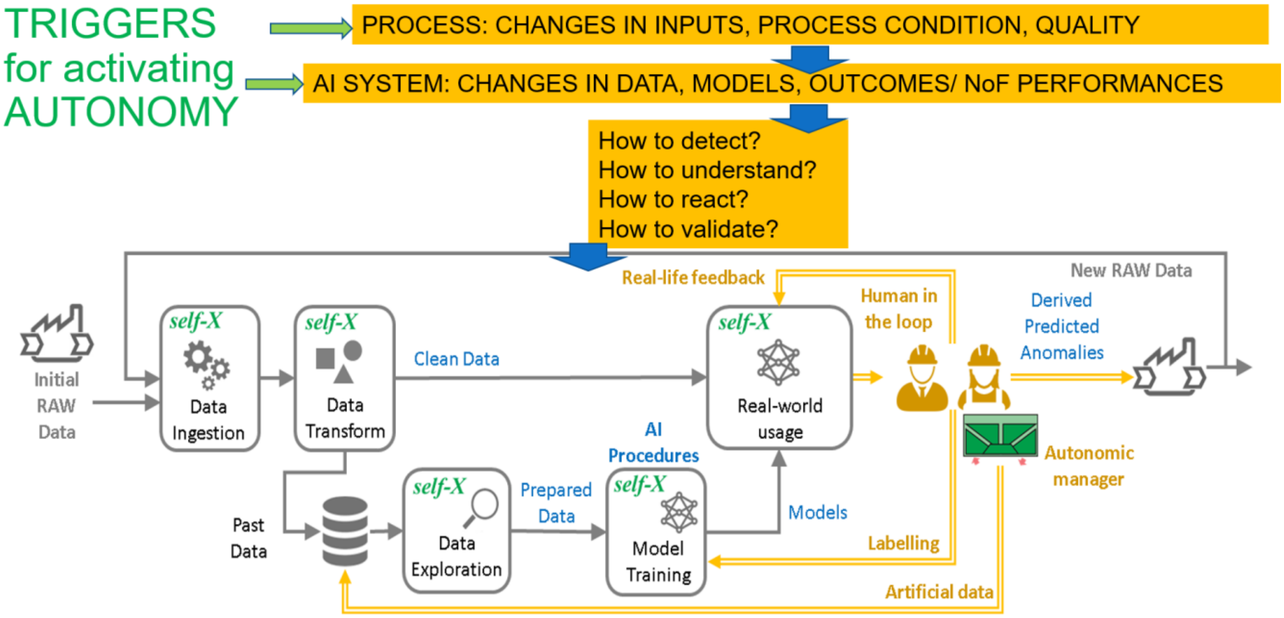


Figure 40. Triggers for activating autonomous behaviour of the AI pipeline

In the following figure, we explain the relation between the improvement in Use cases and the autonomy of the entire system.

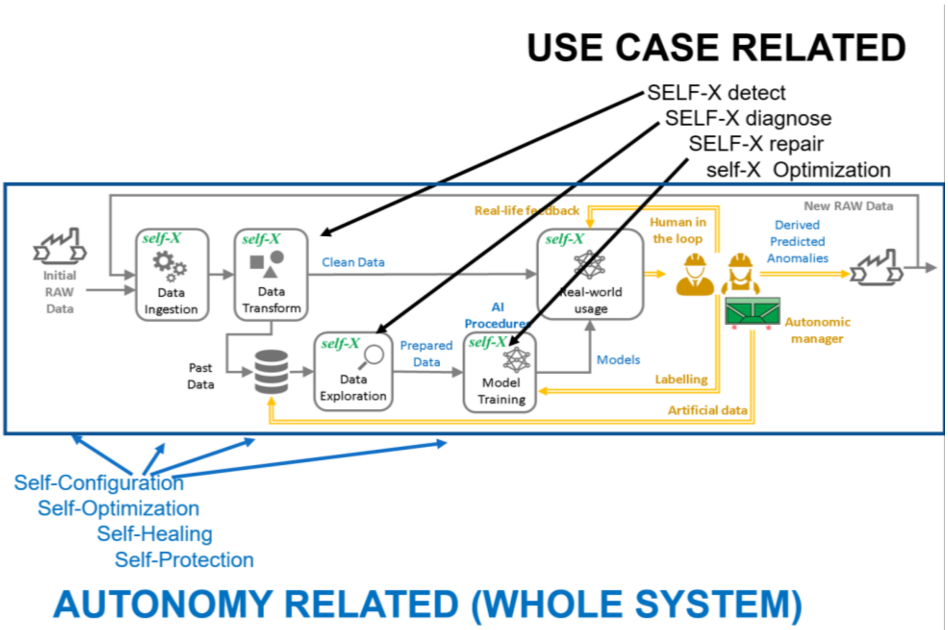


Figure 41. The relation between the improvement in Use cases and the autonomy of the entire system

The way how the requirements from use cases (presented in previous subsection) can be mapped on the AI Data pipeline is illustrated in the following figure.

AM[1] capabilities	SELF-X detect	SELF-X diagnose	SELF-X repair	Feedback for human in the loop	Contributions
Asphalt mix design	of asphalt mix design aggregates and additives' quantities	Asphalt mix according to indicated design		Need of asphalt mix design adaptation	Process, Circularity
Plant elements diagnosis	of asphalt plant performance elements KPI's	source of future failures & need of maintenance		Reports possible maintenance to be performed	Process
Paving conditions and parameters	of adequacy of different mixes paved at job site	possible modifications in the asphalt mix design	Change / adaptation of product mix design	Modifications of current asphalt mix design to adapt to paving site conditions	Process
Quality traceability at lab level	Quality control of final mixes al lab level	Asphalt mix quality adequacy according to standards		Indication of quality parameters according to desired KPI's	Sustainability

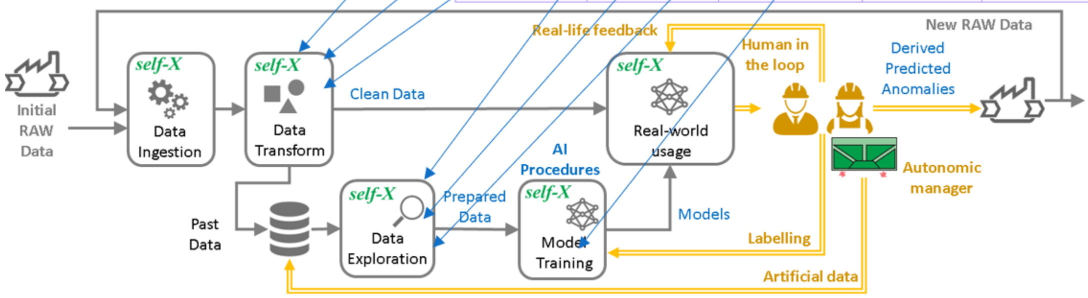


Figure 42. Mapping of the requirements from use cases on the AI pipeline

6 Conclusions

Task T2.1 has been the major technical kick-off of the project as it involved a deep analysis and understanding of the industrial manufacturing processes use cases to fine tune the definition and aim of the different self-X tools. It has been a great challenge as it has been necessary to build a common shared vision among the process industry expert sectors, the researchers looking for technical solutions and the IT experts in the field of automation platform to be able to integrate all the AI procedures.

Besides this, this deliverable explains the generation of the different requirements (functional and non-functional) of self-X abilities for AI procedures and the different components for AI-based applications that are going to be developed for the different use cases. It defines how the different tools will be coordinated (prioritization) of the AI Data pipeline and their interaction (via UI) with the corresponding human role in each use case scenario.

The requirements also make an analysis of the different possibilities available at IT infrastructure level and available components in each use case for the possibilities of the different self-X tools and associated capabilities. That includes the availability of data sources and IT systems to interact with the "Data ingestion" components layer for the self-X AI data pipeline.

Based in all of this, within the requirements, different tentative categorized KPIs have been defined for each of the use cases, divided in: KPIs (socio-environmental-business-technical) affected to later evaluate them in the demonstration phase, later on at the end of WP5.

All these defined scenarios and requirements lays the foundation of the needed extended data pipeline with self-X abilities for each of the AI component of the defined AI Data pipeline in the project, where AI-based solutions will be able to resolve "new and/or unforeseen issues" and adapt over time. The identified data sources (coming from sensors or be data already pre-processed) will be ingested into the pipeline for the Autonomic Manager for AI components to follow the corresponding flow.

In conclusion, although the project has just started the deep exploration done in this task constitutes a solid foundation for the definition and integration of the different described self-X tools of the demonstrators into the Reference Architecture that is being defined in Task 2.2. It will guide the development and refinement for the integration of the components in the AI Data pipeline for human support in WP3 and the corresponding autonomic managers in WP4 respectively.