capri

Cognitive Automation Platform for European PRocess Industry digital transformation

Deliverable

D3.1 CAPRI final reference architecture

Deliverable Lead: Politecnico di Milano

Deliverable due date: 30/09/2021

Actual submission date: 30/09/2021

Version: 1.6

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

Document Control Page	
Title	CAPRI final reference architecture
Lead Beneficiary	Politecnico di Milano
Description	Final architecture of the cognitive automation platform, starting from the three domains cognitive solutions' requirements
Contributors	Silvia Razzetti (POLIMI), Sergio Gusmeroli (POLIMI), Antonio Salis (ENG), Gabriele De Luca (ENG), Mattia Marzano (ENG), Nenad Stojanovic (NISSA), Christoph Nolle (BFI), Spyridon Fourakis (CORE), Jakob Rehrl (RCPE), Ana Margarida Pinto (AIMEN), Zigor Lopez (MSI), Cristina Vega (CAR), Daniel Gomez (CAR), Norbert Holzknecht (BFI)
Creation date	28/07/2021
Type	Report
Language	English
Audience	Public
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

Document History			
Version	Date	Author(s)/ Reviewer(s)	Status
0.1	28/07/2021	Sergio Gusmeroli (POLIMI), Silvia Razzetti (POLIMI), Antonio Salis (ENG)	First ToC
0.11	02/09/2021	Cristina Vega (CAR)	Draft CAR contribution in Ch2
0.12	15/09/2021	Daniel Gomez (CAR), Ana Margarida Pinto (AIMEN)	Some corrections in chapter 2
0.2	16/09/2021	Mattia Marzano (ENG), Gabriele De Luca (ENG), Antonio Salis (ENG)	First contents in chapter 4; updates and extensions in chapter 2; unified baseline

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

1.2	17/09/2021	Christoph Nolle (BFI), Spyridon Fourakis (CORE), Jakob Rehrl (RCPE), Norbert Holzknecht (BFI), Zigor Lopez (MSI)	Contribution from CORE, RCPE, BFI, MSI
1.3	22/09/2021	Mattia Marzano (ENG), Gabriele De Luca (ENG), Jakob Rehrl (RCPE)	Added content in 4.2, updates in 2.3
1.4	27/09/2021	Daniel Gomez (CAR), Spyridon Fourakis (CORE)	Updates in chapter 2 and review by CORE
1.5	29/09/2021	Silvia Razzetti (POLIMI), Nenad Stojanovic (NISSA	Updates in chapter 3 and 5
1.6	30/09/2021	Silvia Razzetti (POLIMI), Antonio Salis (ENG), Laura Rodriguez (MSI), Cristina Vega (CAR)	Includes review from MSI and ENG

List of Acronyms and Abbreviation	
Acronyms	**Description**
AAS	Asset Administration Shell
ACID	Atomicity, Consistency, Isolation, Durability
API	Application Programming Interface
AR	Augmented Reality
BI	Business Intelligence
CS	Cognitive Solution
DA (OPC)	Data Access
DB	Data Base
DLL	Dynamic Link Library
DOE	Design Of Experiment
DT	Digital Twin
ERP	Enterprise Resource Planning
FBD	Fault Based Detection
GDPR	General Data Protection Regulation
GE	Generic Enabler
HDFS	Hadoop Distributed File System
IDS	International Data Space
IDSA	International Data Space Association

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

IT	Information Technology
IoT	Internet of Things
IIoT	Industrial Internet of Things
LD	Linked Data
MES	Manufacturing Execution System
MPC	Model Predictive Control
MQTT	Message Queue Telemetry Transport
OPC	Open Platform Communications
OS	Open Source
PICO	Process Industry COgnitive
PLC	Programmable Logic Control
RA	Reference Architecture
RAP	Reclaimed Asphalt Pavement
RPC	Remote Procedure Call
TSG	Trouble Shooting Guide
UDP	User Datagram Protocol
UA (OPC)	Unified Architecture
UI	User Interface
VR	Virtual Reality
WP	Work Package

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

Table of Contents

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

Table of Figures

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

List of Tables

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

DISCLAIMER

The sole responsibility for the content of this publication lies with the CAPRI project and in no way reflects the views of the European Union.

EXECUTIVE SUMMARY / ABSTRACT SCOPE (POLIMI)

Deliverable D3.1 – "CAPRI final reference architecture" describes the final definition of the CAPRI Reference Architecture (CAPRI RA), as a functional and modular architecture to support innovative cognitive technologies and solutions inside the process industry. It represents the natural sequel of D2.1, that at an early stage of the project described the first definition of the RA, starting from the state of the Art of existing models and standard, that is, from a conceptual point of view. Conversely, D3.1 shapes the **CAPRI Reference Architecture as a result of more than a year of activity in WP3**, validating what it was proposed in D2.1.

Actually, D3.1 is the first document associated to WP3 - "Smart modules for cognitive process industry plants", which is the work-package where the Cognitive Solutions (CSs) are implemented at laboratory level. Since it is expected at Month 18 (after more than a year of activity), it means that the CSs have already taken a shape close to the final solution and this allows to have a large amount of information useful to outline the platform's main components and its architecture, which is the scope of the deliverable. Even if D3.1 is not the one-by-one description of all the activities run in WP3 dealing with the implementation of the 19 CSs, all the results shown in the present document couldn't be achieved without the work done to develop them.

For instance, Section 4 about the "CAPRI Open-Source reference architecture", which is the core of the entire deliverable, provides the description of the main components of the CAPRI Cognitive Automation Platform (CAP) where the CSs will be integrated. It is the result of several meetings involving pilots, technology providers and both WP3 and WP4, in order to collect all the information required to outline the platform. Activities run in WP3, from the Lab hardware/software implementation to the group discussion about the different cognitive aspects of the CSs, are all indirectly reflected in the final reference architecture, here presented.

To be precise, the Cognitive Automation Platform is the result of the **bottom-up approach** chosen to develop WP3, that starts from the three separated domains (Asphalt, Steel and Pharma) to reach a common solution (the CAP) able to fulfil the three sectors' needs and expectations. It means that the three domains maintain a certain independence and the platform can be easily customized to fit their requirements; however, the final architecture consists of a single solution that may be eventually extended and generalized for the process industry. The separation among Asphalt, Steel and Pharma sectors is kept in Section 2 and Section 5 where CAP requirements and result of data analysis respectively are presented separately.

Besides the specificity of the three domains, WP3 deals with **four cognitive components to be implemented in industry** (Smart Industrial IoT connection, Smart Event Processing, Smart Knowledge and Semantic Data Models, Smart Decision Support): the CSs are usually transversal to them, since they encompass more than a cognitive level each and this is why the analysis described in the current document takes into account several different cognitive aspects.

Section 3 about the **implementation of cognitive process** explains how the cognition concepts developed so far works in the 19 CSs in general and in pilots' dataset.

D3.1 is the result of a strong collaboration among the CAPRI partners, who have contributes to outline the main concepts described in the document from a different perspective: the technology providers have updated the list of requirements for the CAP for the integration of the Cognitive Solutions; the system integrator has outlined the final reference architecture; the partners working on pilot datasets have described their main outcomes coming from the running data analysis; and

the integration of the cognitive concepts in the CSs and inside the CAP is described both from an analytical and technical point of view.

1 Introduction

1.1 Scope of Deliverable

Deliverable D3.1 – “CAPRI final reference architecture” aims at providing the methodologically reference implementation of cognitive solutions for the process industry that is derived from the activities carried out in Tasks T3.1 to T3.4.

As it is the first “paper report” of WP3, it represents so far also the main tangible description of WP3 activities, by reporting the updates about the Cognitive Solutions requirements and by detailing the current analysis running about process industry datasets analysis and modelling.

Hence, the main objective is to provide a full picture of the main steps required to implement a cognitive system (tailored on CAPRI pilots requirements) from the platform to the analytics process.

1.2 Audience

D3.1 – “CAPRI final reference architecture”, despite is a public report and it is accessible to anyone who is interested in the topic, it is quite a technical document and it is addressed mainly to a reader with a technical background related to Data Architecture and Dataset analysis.

To properly understand the document, it is required to be familiar with the concepts of architectural component, since the deliverable is a detailed description of the current proposal for CAPRI Open-Source Architecture to finally implement the Cognitive Automation Platform (CAP). In addition, a basic understanding of data analysis and ontology will facilitate the reading.

Inside CAPRI project, D3.1 is conceived to be a report accessible to all partners:

- For the three pilots, it is a technical description of the Cognitive Automation Platform that will be installed in their environment and that will be integrated with existing assets and solutions;
- For technology providers involved in WP4 and WP5, it represents the guidelines to drive the CAP implementation at first stage and its validation at second stage, verifying that it is compliant with pilots requirements;
- The deliverable also represents an inspirational document toward the process industry data analysis, addressed to the partners more familiar with analytics. In this framework, two partners are actively working on pilots’ data analysis and D3.1 is a proper formalization of their activities, useful to share techniques and results.

1.3 Relationship with other deliverables and work-packages

D3.1 is strongly related with D2.1 since it is the conceptual sequel. If the latter is a preliminary design of the CAP Architecture based on existing standards and requirements from use cases that provides a high-level analysis of the state of the art of Open-Source Architecture, on the other side the former describes the final architecture tailored for CAPRI needs, derived from the analysis of the Cognitive Solutions. Under some aspects, D3.1 is an update of D2.1, reflecting the ten months of activity run in WP3 that allow to better set up the platform where the Cognitive Solutions will be integrated.

Deliverable D3.1 will also represent a starting point both for D3.6 and for the entire WP4 and WP5.

D3.6 – “Reference Implementation of Cognitive Process Plants and Alignment with other cognitive initiatives” is expected at month 24 (end of WP3 activities) and it will be the final description of the

This project receives funding in the European Commission’s Horizon 2020 Research Programme under Grant Agreement Number 870062

Cognitive Solutions reference implementation, with the objective of generalising it for process industry. It means, of course, that what is described in D3.1 represents the basis to extend to concept into process industry domain in general.

WP4 – "Cognitive technology solutions for process industry plants" deals with the physical implementation of the CAPRI Cognitive Automation Platform, which is properly described in D3.1. Hence, there is a strong relationship between activities run in WP3 and WP4 and consequently also between deliverables: it is not possible to conceive the two as separated work-packages and deliverables of the first one defines actions of the second one as well.

WP5 – "Prototype demonstrations of Cognitive Automation Platform in CAPRI use cases" will integrate and test the pilots' cognitive technology components inside the Cognitive Automation Platform. Again, D3.1 represents the technical description of the CAP and it will be useful in WP5 since it outlines the main features of the platform and its components fitting with the three domains requirements.

1.4 Document Structure

The document is organized in four main chapters (Section 2 – Section 5), beside the introductive chapter (the current **Section 1**, where the purpose of document, the target audience and the structure are described) and the conclusive one (**Section 6**, summarizing main achievements and addressing future activities).

Section 2 – Reference architecture updates aims at providing an update with respect to D2.1 - Reference architecture of Cognitive Automation Platform. D2.1 was written at the early stage of the project (M8), providing a preliminary picture of CAPRI Reference Architecture based on the existing assets and standards and trying to match them with the three pilots' requirements. Section 5.2 of D2.1 in particular, collects the requirements according to the most updated information at M8. Since some aspects were left on-hold and some other are changed in the meanwhile, our Section 2 will contain, for each domain, an overview of the general architectural approach and Cognitive Solution by Cognitive solution, the main updates with respect to previous deliverable.

Section 3 - Implementing cognitive process in CAP aims at describing how the cognition concepts developed so far works in the 19 Cognitive Solutions in general and in pilots' dataset.
The second objective is to combine the architectural component (fundamental to support a cognitive process built on top of it) with the cognitive process itself, matching the requirements of the CSs into the CAP, that is mapping the cognitive aspect into the platform.

Section 4 – CAPRI Open-Source reference implementation is, hereby, the core of the document since it describes the main features of the CAPRI Open Source Reference Architecture, considering both the horizontal and vertical layers. Then, for each domain, the specific aspects are identified: actually, not all the CAP's blocks are expected to be implemented by each pilot and those related to the specific use-case are highlighted and described in detail.

Section 5 - CAPRI Data and knowledge reference implementation aims at describing the reference implementation from the Data and Knowledge perspective. An entire paragraph is dedicated to data analysis, collecting the results achieved so far by running a preliminary analysis over the three pilots' datasets. For each one, the steps followed to deepen into data are detailed and the first results are shown. The goal of the analysis is to present a structured way to find insights and hidden patterns behind data, taking into account the specificity and differences among Asphalt, Steel and Pharma datasets.

A second paragraph is dedicated to ontologies, with the objective of studying ontologies for process industry, starting from pilots' examples and by validating the applicability of discrete manufacturing ontologies in the domain of process industry.

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

2 Reference Architecture Updates from CSs

2.1 Asphalt use case

2.1.1 General approach

In order to understand the **modular architecture** that supports the Cognitive Solutions implemented in the Asphalt use case, the architecture must be connected with different sources of data from shop floor level (plant level) (e.g., sensors, SCADAs, PLCs).

The current situation of the implementation of this architecture can be seen in Figure 1 (AS-IS situation) whereas the updated (from the ones explained in previous deliverables point-of-view) physical implementation that is being undertaken of this architecture can be seen in Figure 2 (TO-BE situation).

At Gerena Asphalt Plant, main data come from different sensors. There are two types of connectivity to be able to use those data:

A) **Local plant SCADA system developed by plant provider (INTRAME Company) sends PROPRIETARY UDP data-frames** (already used protocol in different EIFFAGE plants) through the local network. In this case, all sensors and actuators **data used for production** (and connected to the plant PLC) are sent through this system. Frames may contain information that is not exclusively sensor data, but "more elaborate" data before being sent into the frame. Previous analysed options were that a local computer called PC CAPRI would be running a local script that “sniffs” local network in order to be able to read and store sensor data according to a specific known mechanism but due to some limitations (“sniffer” program to be used differs from the actual Data Frames that SCADA system is sending due to some updates made by plant provider) this option has been disregarded.

B) **On the other hand, a WAGO (PLC) datalogger has been implemented for the EIFFAGE “Connected Plant**” platform that receives data from another series of sensors, some of them specifically installed and commissioned for CAPRI project and sends them to the central repository of this platform through MQTT protocol. In addition, the WAGO system also reads the data coming from the SCADA in the proprietary UDP data-frame protocol. This WAGO system is able to send all this data to two different IP addresses. This possibility is used to send that same data that is being sent to the EIFFAGE Connected Plant platform to another IP address where a PC running outside of EIFFAGE infrastructure (currently at CARTIF facilities) reads that data that comes using MQTT protocol. This PC is being called **CAPRI server** for the Asphalt use case.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

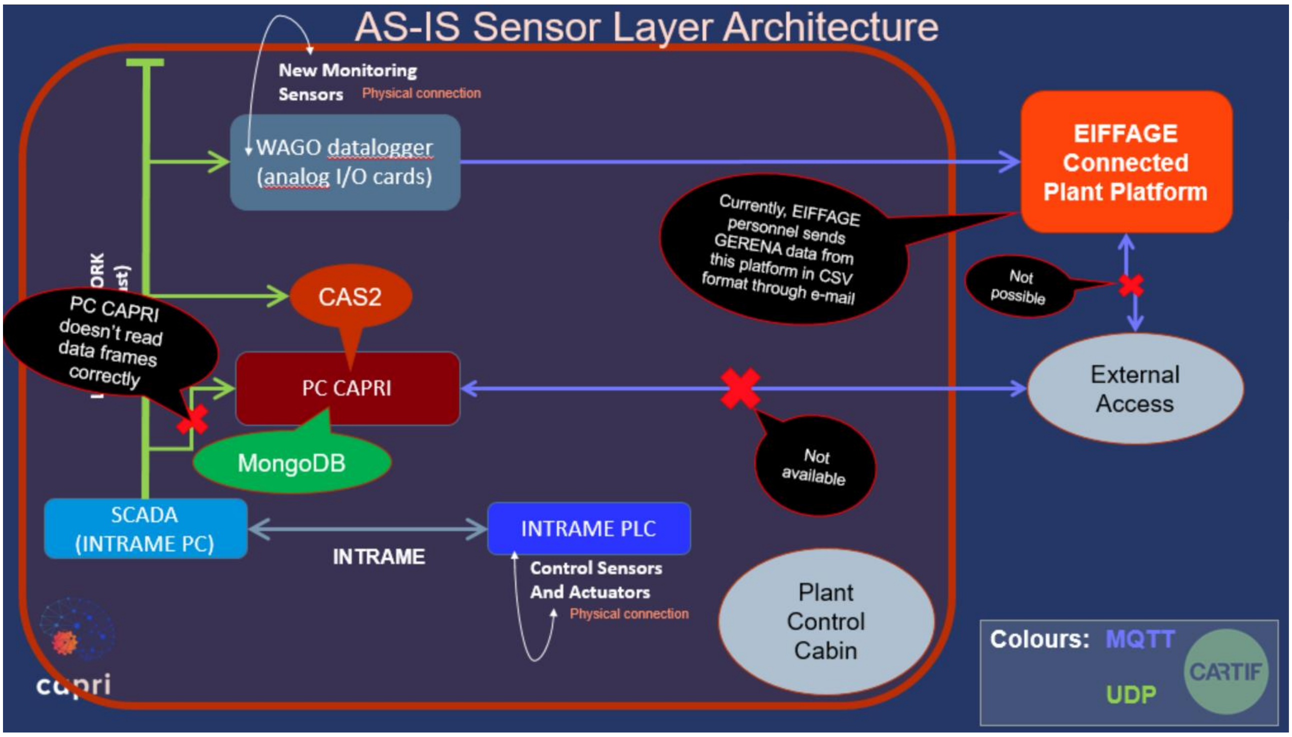

Figure 1 AS-IS Sensor Layer Architecture at Asphalt use case

Current situation (AS IS, Figure 1) of the Gerena Asphalt plant is the following one:

- SCADA system (as previously explained) sends its data-frames through the plant control cabin local network (programmed by plant provider).
- A local PC called PC CAPRI reads these data using a "sniffer" program and stores those data in a local MongoDB database. However, due to some updates done by plant provider to their production control program, that "sniffer" program does NOT read data frames correctly.
- As previously explained, there is a WAGO PLC datalogger which reads all those data-frames and along with data from several directly connected sensors to that system, they are sent to the EIFFAGE Connected Plant platform (a company monitoring platform for all their company plants).
  - This could be used as a central point to access Gerena plant data but due to some restrictions (company privacy policies, etc), external access from third parties is not possible.
  - Another possible data access point could be PC CAPRI, but external access to this computer is not available as well.
- In addition, several data coming from different sensors commissioned at the asphalt plant by CARTIF personnel is used for the development of CAS2 Cognitive Solution. Several programs locally running at PC CAPRI are in charge of gathering and analysing those data.
- Due to all of this, for CS's development, current Gerena sensors data is sent by EIFFAGE personnel through e-mail in *.csv files and uploaded to the project central repository accessed by all project partners.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

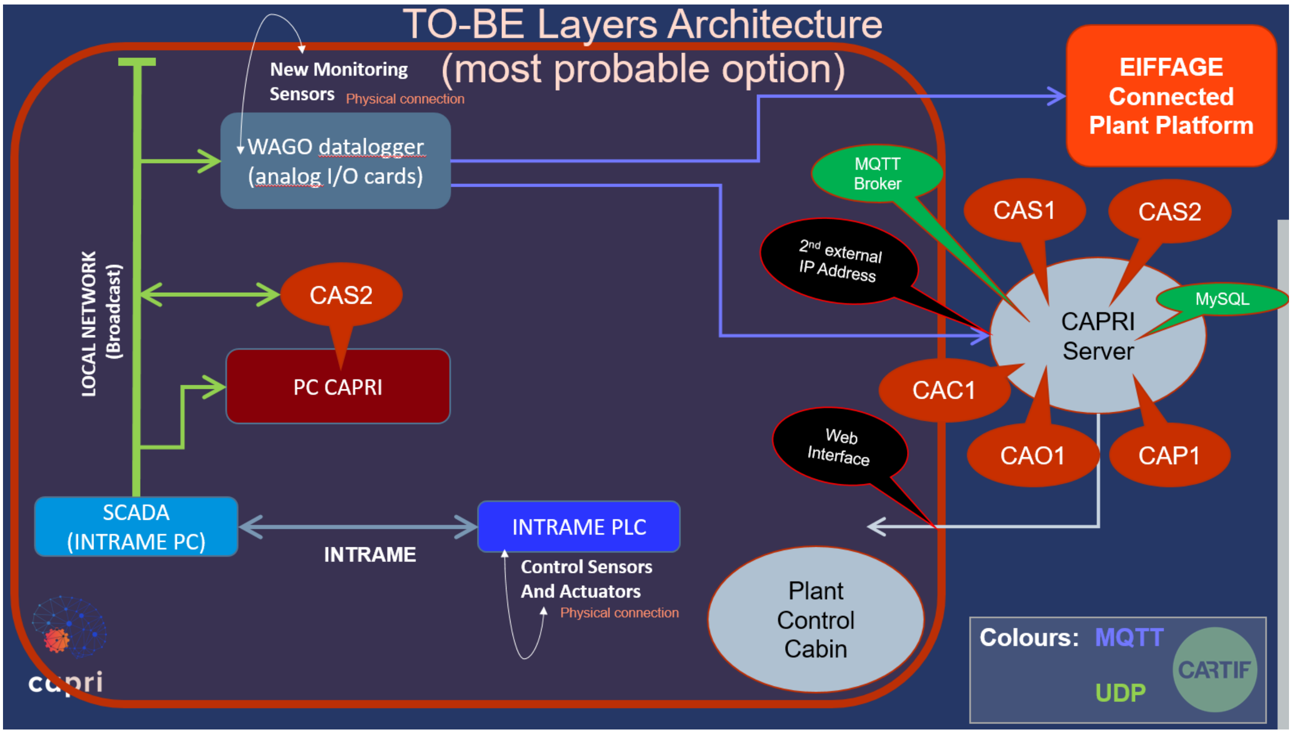

Figure 2 TO-BE Layers Architecture in Asphalt use case

As previously explained, in the TO BE scenario (Figure 2) at plant control cabin there is a local PC called **PC CAPRI** which is connected to the control cabin LOCAL NETWORK. Currently, this PC runs software that gathers data coming from sensors regarding CAS2 cognitive solution.

In this case, the mentioned **WAGO PLC datalogger** runs as **a central point of data gathering** (both production data sent using the explained UPD data-frames coming from sensors and actuators connected to the PLC and all those sensors directly connected to the WAGO system).

Using the possibility of sending data to 2 different IP's from the WAGO system has been undertaken due to some limitations to connect to the EIFFAGE Connected Plant platform from third-parties outside of the EIFFAGE consortium (privacy policies and so on). The **CAPRI server** is able to gather all data coming from the plant thanks to the WAGO system and it is stored in a local database (at the server side). To do that a locally installed MQTT Broker reads all coming data from the WAGO datalogger and thanks to the ORION Context Broker and the persistence layer, also locally installed at the CAPRI Server, data is stored in the mentioned DB. The different cognitive solutions that are currently being implemented run in this CAPRI server:

- CAC1 cognitive solution runs in the CAPRI Server (according to its algorithm).
- CAS2 cognitive solution runs in the CAPRI Server (according to its algorithm).
- CAO1 cognitive solution runs in the CAPRI Server (according to its algorithm).
- CAP1 cognitive solution runs in the CAPRI Server (according to its algorithm).
- CAS1 cognitive solution runs on its OWN, with SEPARATE hardware and software system. Nevertheless, it will be connected to the CAPRI server in order to read the information needed from sensors in the plant and write the corresponding output data.

OUTPUTS of the different CS's will be displayed through a **Web Interface** that can be accessed at the plant control cabin giving the corresponding outcomes to the plant operators or managers as

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

a first step for testing the different CS's. The different alarms, warnings and data streams of the COGNITIVE SOLUTIONS will be shown through this web interface.

These solutions running at CAPRI server side will run on a cloud-based solution. Most outcomes to be applied by the different solutions can run in this approach due to the fact that the asphalt mixing process is a process where the different main variables (temperature, flow, etc.) have a settling time in the order of several seconds or minutes. This means that "almost" real-time access (in the order of milliseconds or just one second) to local resources is not mandatory and it does not imply a restriction to be considered in order to deploy an edge solution. The described architecture to access and show the outcomes of the different Cognitive Solutions is sufficiently robust to be commissioned for this use case (the asphalt mixing plant).

The different mapping of the Asphalt use case CAPRI cognitive solutions to the Reference Architecture is explained in the following points only reflecting the changes done with respect to the ones stated in D2.1 deliverable:

2.1.2 (CAS1) Cognitive Sensor of Bitumen Content in RAP (Reclaimed Asphalt Pavement)

- ✓ **IIoT –** No new updates
- ✓ **Data in Motion** – No new updates
- ✓ **Situational Data –** No new updates
- ✓ **Smart Data Space and Application –** It will provide output of sensor (%bitumen in RAP) to CAP for optimization of the asphalt mix formula. Since several parameters are considered in the asphalt mix formula, being the % of bitumen in RAP one of them, this optimization cannot be done by CAS1 directly.
- ✓ **Data at rest** – No new updates
- ✓ **System Wrappers** – No new updates
- ✓ **Authorization and Authentication** – No new updates
- ✓ **Identity Management** – No new updates
- ✓ **Data Protection and Privacy** – No new updates

2.1.3 (CAS2) Cognitive Sensor for Filler Measurement

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application –** Part of the cognitive algorithm runs through the edge node called PC CAPRI and part in the cloud (CAPRI Server)
- ✓ **Data at rest –** Currently using *.csv files. However, static data coming from CAPRI Server DB once commissioned.
- ✓ **System Wrappers –** Currently using *.csv files but historical data coming from the mentioned CAPRI Server DB once commissioned.
- ✓ **Authorization and Authentication –** No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – no new update

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

2.1.4 (CAC1) Cognitive Solution for Optimal Control of Dryer Drum

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – Cognitive algorithm for the control of dryer drum through cloud node (CAPRI Server). Even though it is a control algorithm, there are no real-time constrains due to long settling and steady state times (controlled variables are long time evolution temperatures, no need to run in milliseconds or seconds time constrains but minutes).
- ✓ **Data at rest** – Currently using *.csv files. However, static data coming from CAPRI Server DB once commissioned.
- ✓ **System Wrappers** – Currently using *.csv files but historical data coming from the mentioned CAPRI Server DB once commissioned.
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.1.5 (CAO1) Cognitive Solution for Predictive Maintenance of Baghouse Filter

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – Cognitive algorithm for the predictive maintenance of the baghouse filter running at CAPRI Server (cloud node)
- ✓ **Data at rest** – Currently using *.csv files. However, static data coming from CAPRI Server DB once commissioned.
- ✓ **System Wrappers** – Currently using *.csv files but historical data coming from the mentioned CAPRI Server DB once commissioned
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.1.6 (CAP1) Cognitive Solution of Planning and Optimization of Asphalt Production

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – Management and aggregation of context data (pressure, temperature, humidity, etc.)
- ✓ **Smart Data Space and Application** – Cognitive algorithm for the generation of the settings of the set points of the controllers to optimize production running at CAPRI Server (cloud node)
- ✓ **Data at rest** – Currently using *.csv files. However, static data coming from CAPRI Server DB once commissioned
- ✓ **System Wrappers** – Currently using *.csv files but historical data coming from the mentioned CAPRI Server DB once commissioned
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

2.2 Steel use case

2.2.1 General approach

In the steel use case data from the SCADA (Level 2) and MES (Level 3) systems in the steel plant needs to be accessed by the different cognitive solutions. Data comes from different processing steps, in particular secondary metallurgy, continuous casting, hot rolling and one of the finishing lines (Figure 3). The product tracking system for CSS1, consisting of two cameras and two painting devices, is connected to the MES, hence no special treatment of this system is required. The CSS2, CSS3 and CSS4 are calculating unmeasurable process information and CSS5 finally uses all outputs from the other CSS to estimate the appearance of product quality anomalies.

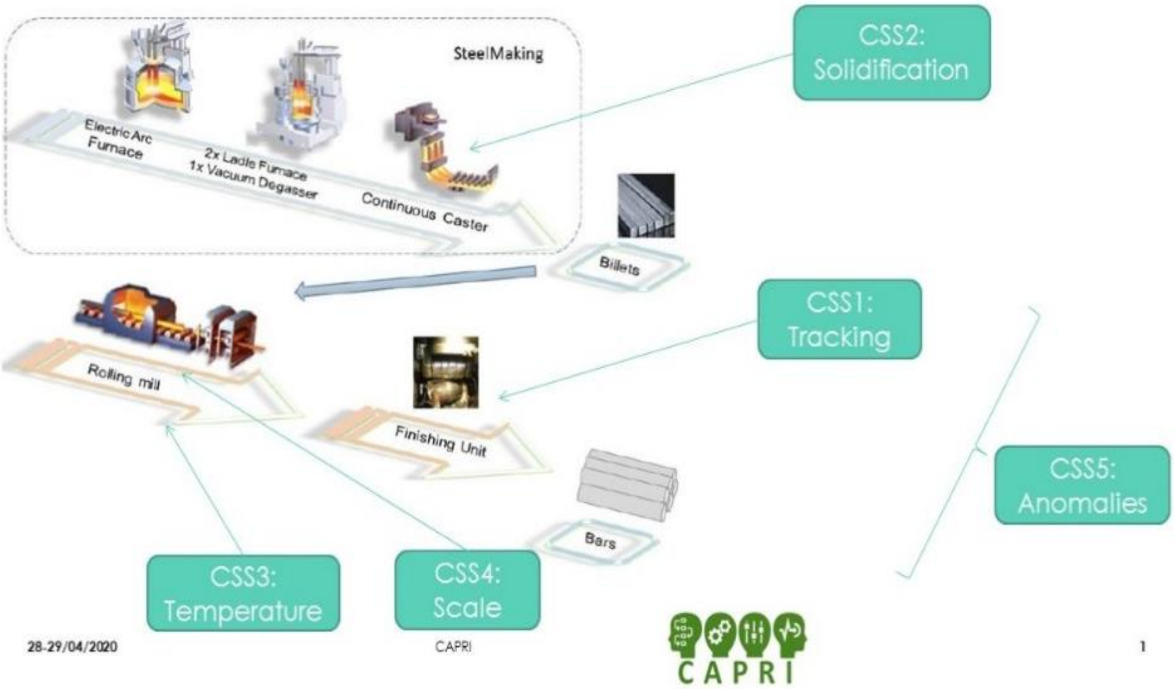

Figure 3 Overview of the steel use case processing steps and cognitive solutions

The AS-IS situation regarding data access from the automation system involves a manual CSV export by SIDENOR staff and transfer via email. The data is then imported into a database and a data access service has been set up which provides access to the data via a web API. The digital twin platform retrieves the data via this interface and stores it in its own database. The cognitive sensors can then query the data via the digital twin service. As an alternative route it is possible to import the exported CSV files directly to the digital twin database, as shown in **¡Error! No se encuentra el origen de la referencia.**.

For the TO-BE architecture the manual CSV export/import and the redundant data storage will be removed. Data ingestion from the plant automation system to the CAP will be taken care of by dedicated agents pushing the data to a data broker, and storage will be available exclusively within the CAP. Furthermore, the CAP will provide an interface for online data access, streaming data directly to the client as it arrives. This is illustrated in **¡Error! No se encuentra el origen de la referencia.**.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

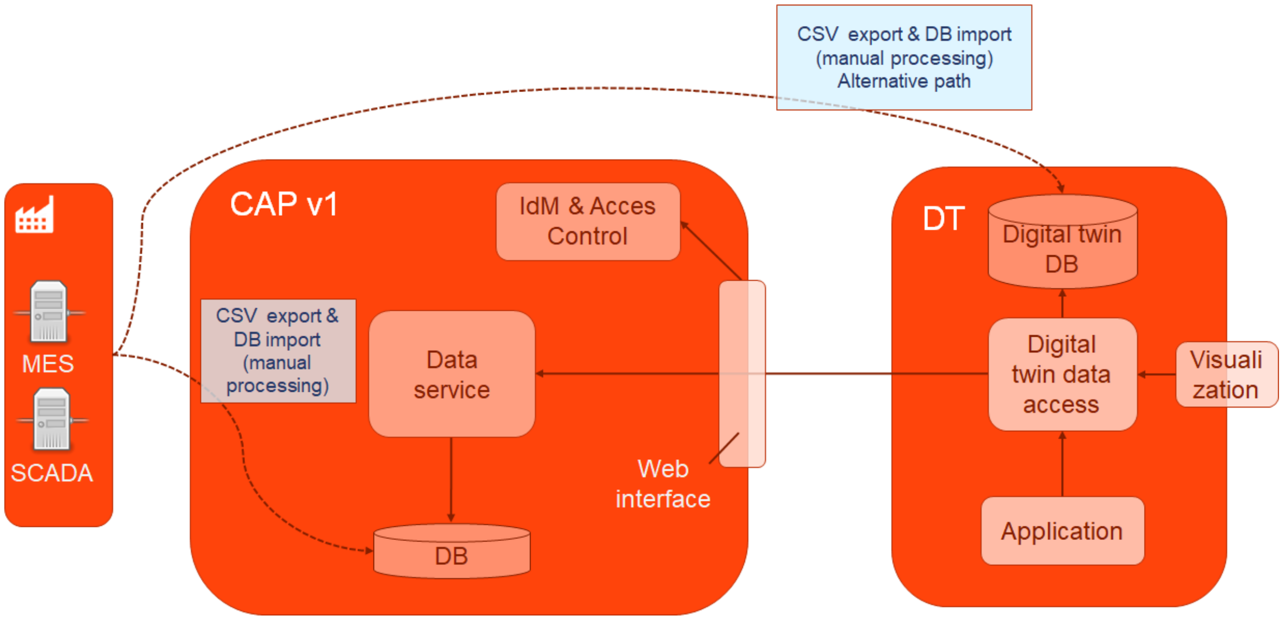

Figure 4 AS-IS sensor layer architecture steel use case. CAP v1 denotes the preliminary cognitive automation platform; DT is the digital twin platform hosting the cognitive solutions (applications).

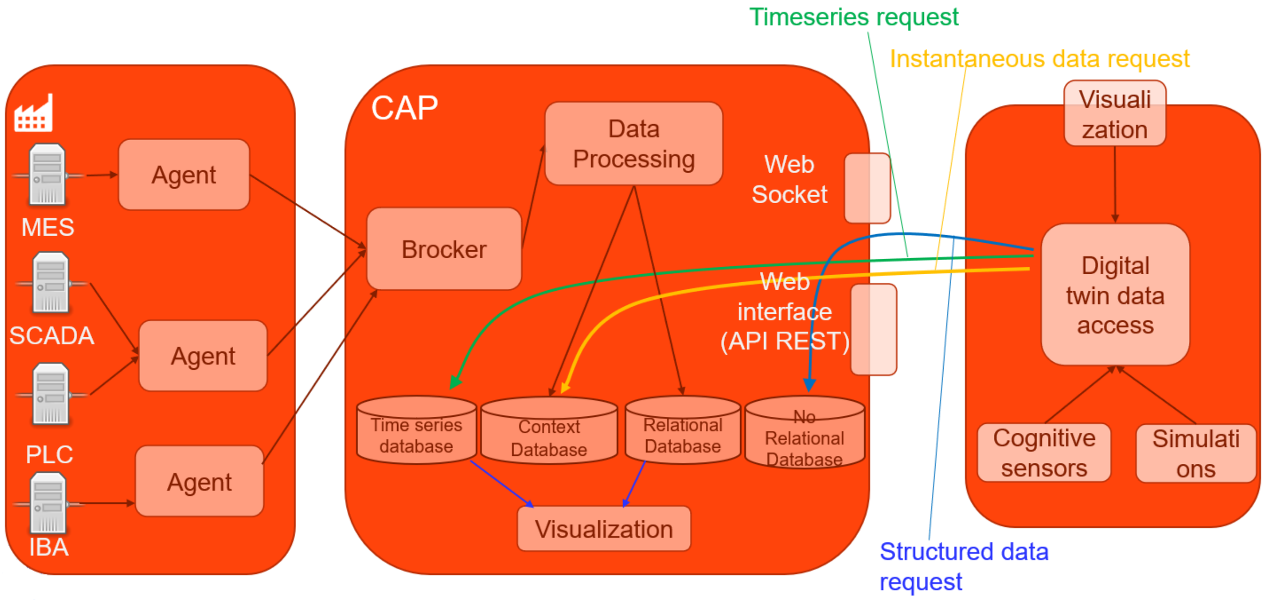

Figure 5 TO-BE sensor layer architecture steel use case. CAP denotes the Cognitive automation platform; DT is the digital twin platform hosting the cognitive solutions (applications).

The implementation of the Sensor Layer will be based on existing proprietary components:

- **Agents** are responsible for connecting to the data sources and supplying the data message to the broker for further processing. Agents provide interoperability.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

- **Broker**: Is responsible for collecting (via MQTT) and routing the data messages.
- **DataProcessing**: Processes the data, enriches it if necessary and distributes it to each of the containers with databases.
- **Web interface**: The web interface will be double, API REST for data requests, and web sockets for subscription. The data model will be based on existing ontologies where possible.
- **Databases**: Different standard Open-Source databases are used (time series, context, SQL and No-SQL). Each retrieved data shall be stored in the database type that best fits its structure
- **Visualization**: Visualization is realised with Grafana for Dashboarding and a BI tool for Analytics (SuperSet). Both opensource

2.2.2 (CSS1) Cognitive Sensor for steel products tracking

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the MES (Level 3) and streamed to the CAP broker by means of the dedicated MES agent
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – The currently available dataset for development has been provided in the form of CSV files. In the future, historical data will be stored in a time series database that is part of the CAP storage layer.
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.2.3 (CSS2) Cognitive solidification sensor

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the MES (Level 3) and streamed to the CAP broker by means of the dedicated MES agent
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – The currently available dataset for development has been provided in the form of CSV files. In the future, historical data will be stored partly in a time series database and partly in a structured (NoSQL) database which will be part of the CAP storage layer.
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.2.4 (CSS3) Cognitive temperature sensor

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the MES (Level 3) and the IBA system and streamed to the CAP broker by means of dedicated agents
- ✓ **Situational Data** – No new update

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – The currently available dataset for development has been provided in the form of CSV files. In the future, historical data will be stored in a time series database which will be part of the CAP storage layer
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.2.5 (CSS4) Cognitive scale sensor

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the MES (Level 3) and the IBA system and streamed to the CAP broker by means of dedicated agents
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – The currently available dataset for development has been provided in the form of CSV files. In the future, historical data will be stored in a time series database which will be part of the CAP storage layer
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.2.6 (CSS5) Risk and anomalies sensor

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the CAP (online data access), including process data from the MES and results of the other cognitive sensors
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – Due to temporary non-availability of bar tracking data the risk sensor concept is being adapted. In order to nevertheless be able to connect the quality information of the end products to the process data we decided to consider customer orders consisting of single billets only in the first step. Since the order information is already tracked through the process chain, this implies that we will be able to associate quality information to individual billets (the intermediate products), but not to bars (the final products) as originally envisioned. The risk sensor in its first version will determine a risk estimate for individual billets after the casting process and the hot rolling. After recommissioning of the bar tracking system, it is foreseen to enhance the risk sensor accordingly.
- ✓ **Data at rest** – Historic data for training purposes will be obtained from the CAP (structured data and timeseries)
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

2.2.7 (CSO1) Digital Twins

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – Data will be obtained from the MES (Level 3) and streamed to the CAP broker by means of the dedicated MES agent
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – The currently available dataset for development has been provided in the form of CSV files. In the future, historical data will be stored partly in a time series database and partly in a structured (NoSQL) database which will be part of the CAP storage layer
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3 Pharma use case

2.3.1 General approach

Most of the developed cognitive solutions of the pharma use case are relying on real-time data originating from SCADA systems or specific process equipment. Therefore, the robust interconnection of the solutions is of high importance. The use of industrially established communication protocols is pursued.

¡Error! No se encuentra el origen de la referencia. gives an overview of the cognitive solutions being developed for the pharma use case. There is no distinction on hardware and software components made in this figure. These details can be found in **¡Error! No se encuentra el origen de la referencia.** (AS-IS situation) and **¡Error! No se encuentra el origen de la referencia.** (TO-BE situation). In **¡Error! No se encuentra el origen de la referencia.**, also the current status of the

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

implementation work is indicated. The main idea is to make all the real time data available via OPC UA.

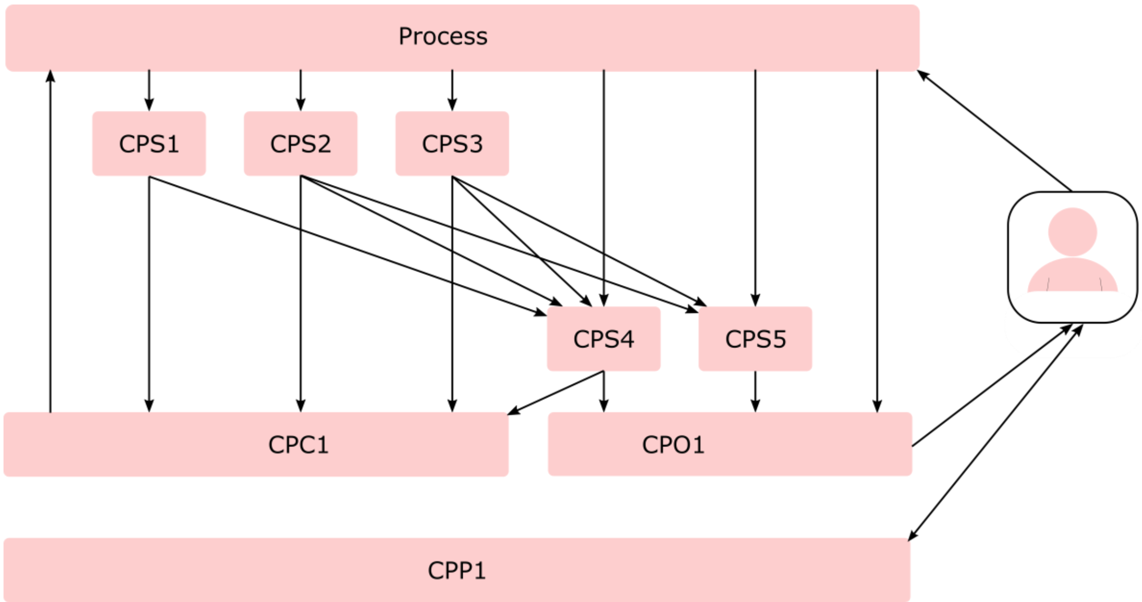

Figure 6 Overview of cognitive solutions

The development of the sensor solutions CPS1 and CPS2 includes both the integration of new sensors to the process, as well as the creation and implementation of respective algorithms. As indicated in Figure 7, CPS1 involves the hardware implementation of a Raman probe to the manufacturing process, as well as the development and implementation of an algorithm which needs to access the raw data from the Raman probe in order to compute meaningful characteristics. The communication is established via OPC UA. Therefore, an OPC UA server has been implemented in Python. Successful tests on a surrogate system have been executed, the implementation on the target system is ongoing. CPS2 covers the implementation of a particle sensor (Parsum Probe) and the development of the respective algorithm for computing the relevant particle properties from the provided raw data. Again, OPC UA will be used for data transfer. Due to the required compatibility with the existing SIPAT server, which is collecting process data, and which only supports OPC DA, a custom built “OPC DA client to OPC UA server” is currently being implemented via Python. The algorithms have been developed, their implementation on the target hardware is ongoing.

This project receives funding in the European Commission’s Horizon 2020 Research Programme under Grant Agreement Number 870062

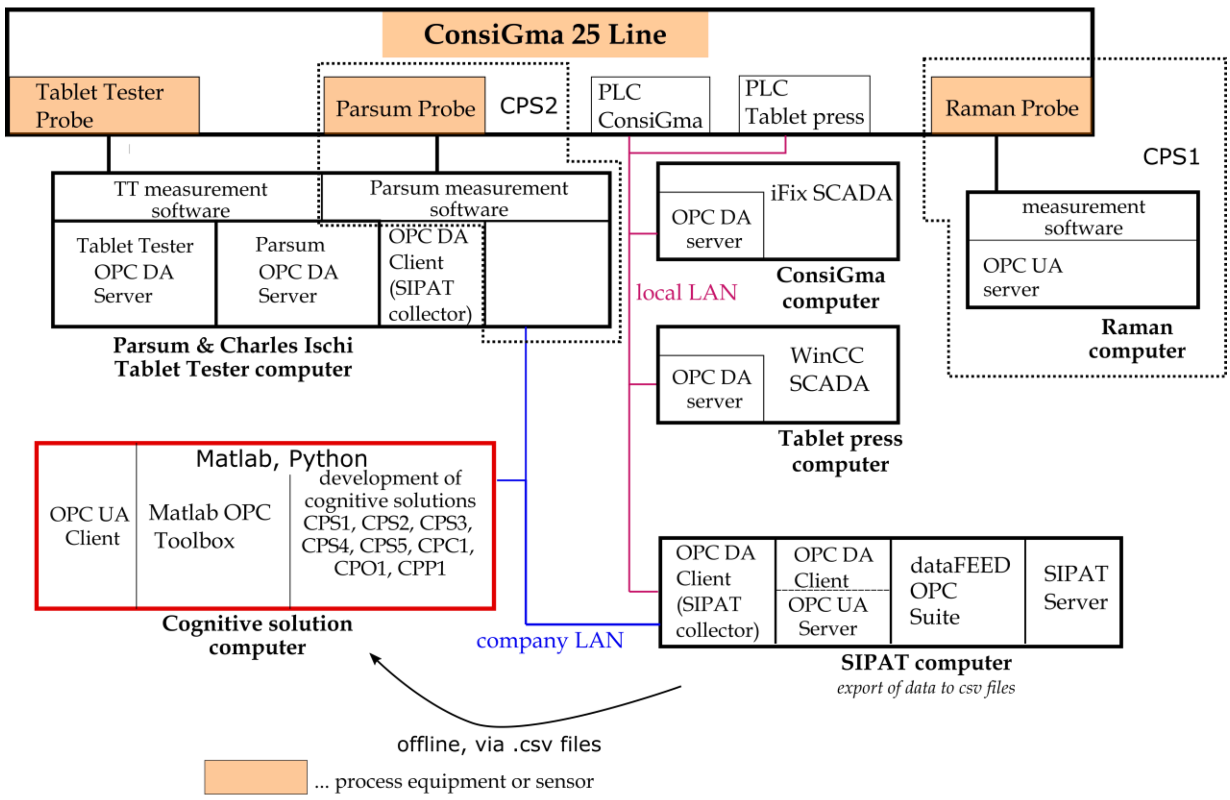

Figure 7 Topology of the pharma use case, AS-IS situation

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

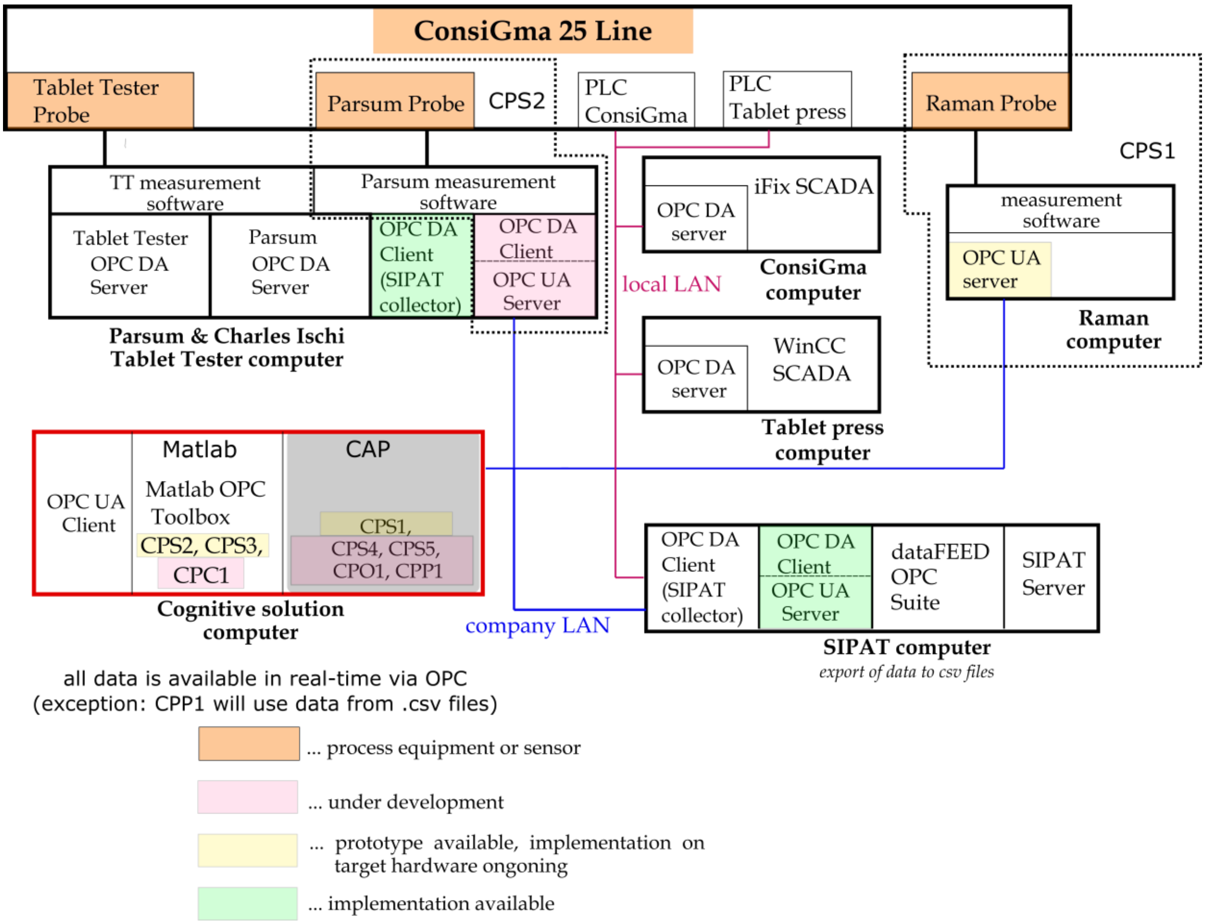

Figure 8 Topology of the pharma use case. TO-BE situation

The sensors CPS3-CPS5 are all either taking data from other CPSx, which are available via OPC UA, or from the process line. The process line data is only available via OPC DA. Consequently, an "OPC DA client to OPC UA server" was implemented in Python, which is running on the SIPAT computer. A CPS3 algorithm is available, a test on the target hardware is planned as a next step. CPS4 and CPS5 algorithms are under development. Part of CPC1 has been implemented already and successfully tested in offline simulations, the implementation on the target hardware and the tests using online data are currently ongoing. CPO1 will also rely on online-data captured via OPC UA and is currently under development. CPP1 will be based on offline data. Its development has not yet started.

AS-IS situation: All data except that of CPS1, CPS2 and the tablet tester are available via OPCUA. Development of the cognitive solution is done on a dedicated computer ("cognitive solution computer"). Data of CPS2 is exported to .csv files via the SIPAT computer and used for offline development purposes at the moment.

TO-BE situation: The OPCUA interface of CPS1, CPS2 and the tablet tester will deliver the respective data in real-time. The CAP will be set up on the cognitive solution computer and some of the cognitive solutions will be implemented via the CAP, some of them will be run in Matlab and appropriate interfaces to the CAP will be in place, see Figure 8.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

2.3.2 (CPS1) Blend Uniformity

- ✓ **IIoT** – Collection of data coming from a Raman probe and no more from other additional sensors and other computing nodes (Consigma, SIPAT)
- ✓ **Data in Motion** – No new updated
- ✓ **Situational Data** – Management and aggregation of context data (Temperature, powder mass flow, liquid mass flow, etc.)
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.3 (CPS2) Granule Quality

- ✓ **IIoT** – Collection of data coming from and Parsum probe and no more from other additional sensors and other computing nodes (Consigma, SIPAT)
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – no new update
- ✓ **Smart Data Space and Application** – Cognitive algorithm computing a characteristic granule size descriptor from particle size raw data (according with real-time constraints this could be provided through edge nodes instead of the cloud).
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.4 (CPS3) Granule Moisture

- ✓ **IIoT** – Collection of data coming from various sensors and other computing nodes (Consigma)
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – Nno new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.5 (CPS4) Dissolution Prediction

- ✓ **IIoT** – Collection of data coming from various sensors, other computing nodes (Consigma, Parsum OR Raman, tablet tester) and cognitive (CPS1-CPS3) sensors.
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update

- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.6 (CPS5) Fault Detection

- ✓ **IIoT** – Collection of data coming from various sensors, other computing nodes (Consigma, Parsum, Raman, tablet tester) and cognitive (CPS2-CPS3) sensors
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.7 (CPC1) Cognitive Control Concept

- ✓ **IIoT** – Collection of data coming from various process sensors, and cognitive (CPS1-CPS4) sensors. Computed values need to be written to actuators
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.8 (CPO1) Cognitive Operation Solution

- ✓ **IIoT** – No new update
- ✓ **Data in Motion** – No new update
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update
- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management** – No new update
- ✓ **Data Protection and Privacy** – No new update

2.3.9 (CPP1) Cognitive Planning Solution

- ✓ **IIoT** – Collection of data coming from various cognitive sensors (no real time data, but data captured during previous trials will be used, SIPAT data will be exported for that purpose)
- ✓ **Data in Motion** – Historic operational data (no real time data, but data captured during previous trials will be used)
- ✓ **Situational Data** – No new update
- ✓ **Smart Data Space and Application** – No new update
- ✓ **Data at rest** – No new update

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

- ✓ **System Wrappers** – No new update
- ✓ **Authorization and Authentication** – No new update
- ✓ **Identity Management –** No new update

Data Protection and Privacy – No new update

2.4 General updates in the Reference Architecture in the CAP platform

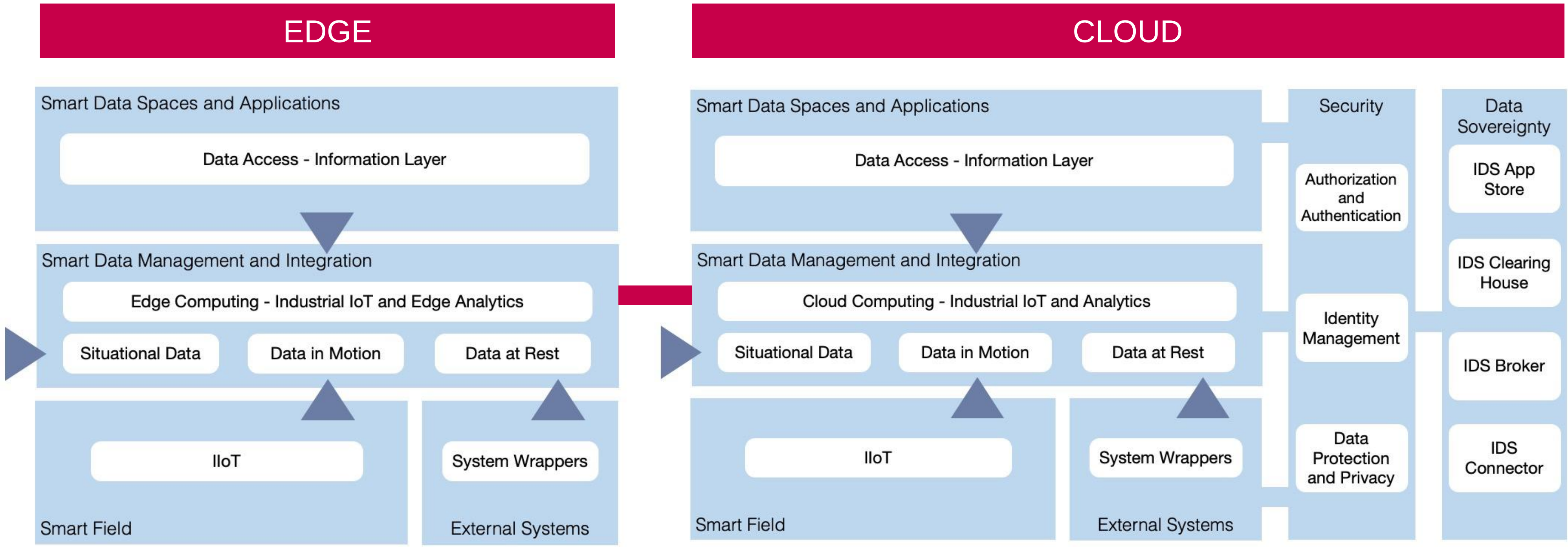

Figure 9 CAP Edge-Cloud Reference Architecture

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

The Reference Architecture of CAP platform (Figure 9) considers the exploration of the different cognitive solutions defined, developed, and integrated in each of the use cases. It follows what is described in *CAPRI D2.2 Use case requirements for cognitive technologies applications in use cases*, where is defined a complete first definition of the requirements for each of the different cognitive solution.

The development of advanced cognitive software solutions is very challenging, and it becomes even more challenging to implement this software in real industrial environments. For this reason, it was done an iterative process, started in *CAPRI D2.1 Reference architecture of Cognitive Automation Platform*, from which there were collected the inputs, the functional and non-functional requirements coming and all proposed cognitive solutions, was obtained the validation from the pilots Asphalt, Steel and Pharma with the ambition to introduce the cognitive automation processes in the process industry.

The main improvement related to the RA is the highlighting of the concept of edge and cloud cognitive computing with the goal of solving business challenges, creating new value from data, and improving the product quality. The implementation hopefully could be done reusing already existing components in the factory, already in place at different levels like data source connectivity level, application logic level with or without historical database, or graphical visualization level.

It is taken in consideration the fact that today a manufacturing company has the chance, in a scenario like the one described in CAPRI, to choose between a Cloud Computing service or investing in an approach such as the Edge Computing technologies. Both technologies are similar regarding the method of storing and processing data, as showed in the picture above, however the difference is strictly related to the physical location of storing and processing, the amount of data, the processing/response time, and other factors.

Obviously, it doesn't exist an absolute choice, but it is strictly related to the needs a company has. If it is needed a high computing power, or in case of big amount of data to process or storage, it should be better to use a Cloud computing service provided by third-party, despite the risk in term of security.

On the other side, Edge Computing could be helpful in case of data generated locally, inside the manufactory company, during the production processes, when for instance it is required a real time processing of data. In that case it is feasible to store, process and analyse acquired data in place.

Usually in cognitive manufacturing solution are gathered data in real time and applied analytics to obtain insights into the process, in this case the Edge Computing has a fundamental role in this scenario where the data from data can be discovered or not pattern, or in any case relevant information in real time. Today the reduced cost of chipsets, sensors and storage can easily allow that, even if another option could be the improvement of data quality on the edge side and then leverage the power of cloud computing.

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

3 Implementing Cognitive Process Plants in CAP

This chapter aims at describing how the cognition concepts developed so far works in the CSs in general, and pilots' dataset.

3.1 Motivation

Challenge for Cognitive Architecture

One of the most important challenges for developing innovative solutions in the process industry is the complexity, instability and unpredictability of the processes, since they are usually running in harsh condition, dynamically changing the values of process parameters, missing a consistent or at all monitoring/measurement of some parameters important for analysing process behaviour.

For cognition-based (cognitive) solutions these are even more critical constraints, since cognition requires (usually) a huge amount of high quality data for ensuring the quality of the learning process (precision, efficiency). Moreover, getting high quality data usually requires an intensive involvement of human experts in curating (or even creating) the data in a time consuming process. In addition, a supervised learning process requires labelling/classifying the training examples by domain experts, which makes a cognitive solution quite expensive.

It is important to emphasize that the role of human is critical, which can be illustrated using the following comparison:

The difference between a **highly experienced expert** and a **highly-talented novice**: the experts have the **sixth sense** for detecting (early enough) variations/unusualities in the process and to **decide** on spot if the unusuality is something that should be followed closer or is just a temporary disruption. And they are usually very reliable in their decisions. In a very competitive industry environment, this skill can be of critical value for optimizing quality control process, e.g. by having a very efficient anomaly detection mechanism.

Process Industry COgnitive PICO architecture

PICO architecture is based on the recent work in cognitive science[1] and reflects the need for explaining the intelligent behaviour through cognitive processes, which are well structured processes in human beings. Following figure illustrates the original cognitive architecture.

[1] Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A Standard Model of the Mind: Toward a Common Computational Framework across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics. AI Magazine, 38(4), 13-26. https://doi.org/10.1609/aimag.v38i4.2744

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

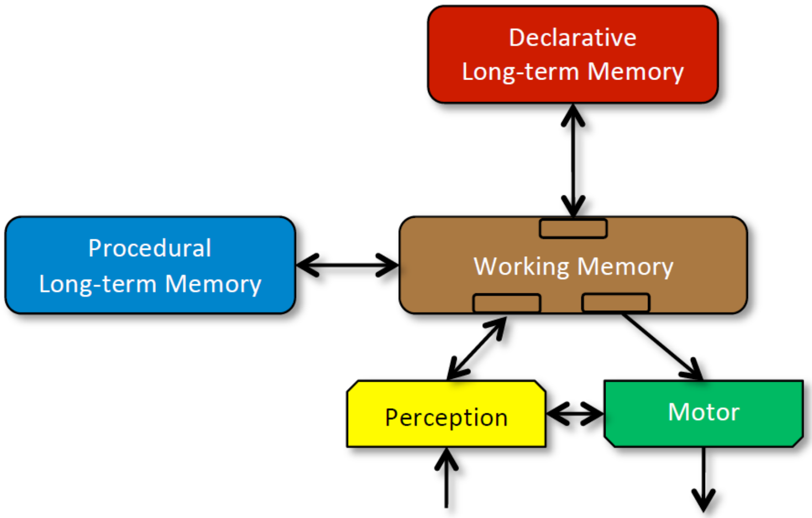

Figure 10 (Cognitive Architecture) across Artificial Intelligence, Cognitive Science, Neuroscience, and Robotics

A cognitive architecture is analogous to a **building architecture**, which describes its fixed structure (e.g., floors, rooms, and doors), but not its replaceable elements (e.g., tables, chairs, and people).

PICO postulates that these human-oriented processes can be mapped into industry processes, like:

- **Cognitive Perception** using dynamic data-driven scanning of the complex industry assets, realizing an efficient, cognition-based collection of heterogeneous data (all senses).
- **Fast Thinking**, enabling an efficient edge-based variations detection and fast understanding of new situations, creating attention for complex, so called unknown unknown situations.
- **Slow Thinking**, enabling complex and efficient processing of complex situations, by creating digital models of their behaviour from sensed data, supporting timely and precise decision making.

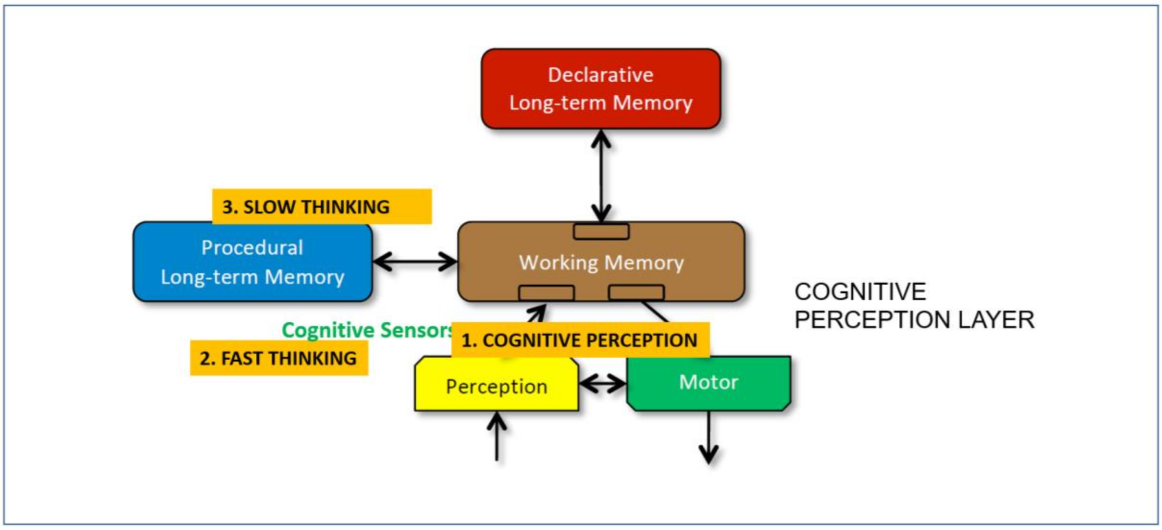

Figure 11 PICO architecture

This project receives funding in the European Commission's Horizon 2020
Research Programme under Grant Agreement Number 870062

3.2 Mapping to CAP

In this section we provide a mapping between the PICO architecture and CAP.

Following figure provides a kind of big picture for that mapping.

Very briefly, **Cognitive Automation Platform** can be analyzed in two dimensions:

- **Smart Cognitive components and**
- **Levels of cognitive human-machine interaction**

Each of them consists of several components, a depicted in Figure 12:

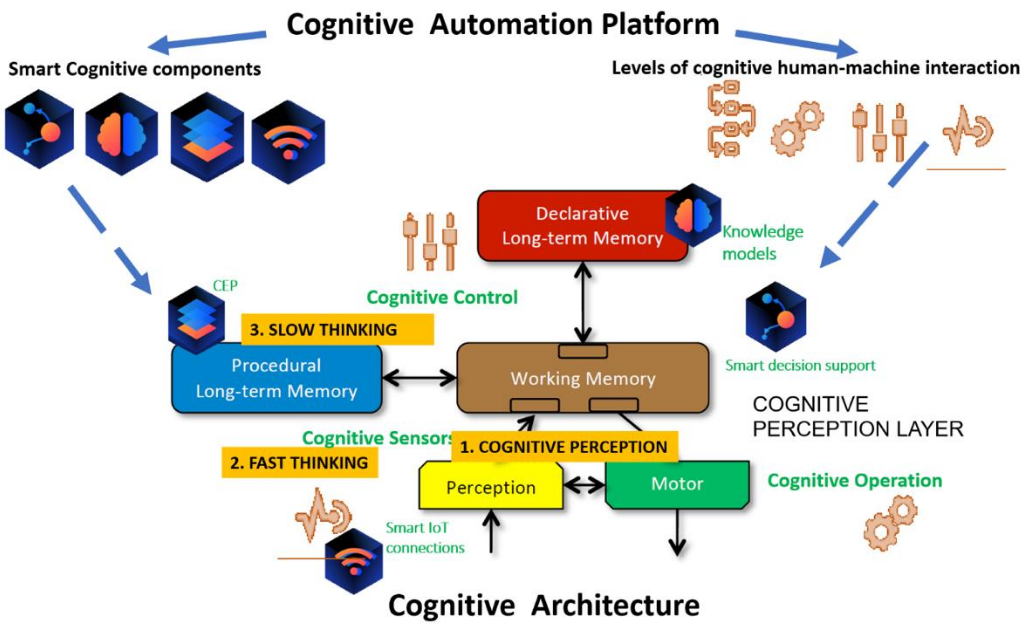

Figure 12 Mapping between CAP and PICO

The Reference Architecture has been integrated (Figure 13**¡Error! No se encuentra el origen de la referencia.**) with those dimensions: the Smart Cognitive components are described by the Processing Layer since, integrating cognitive modules, enables the cognitive perception. On the other hand, levels of cognitive human-machine interaction can be represented by the Application Layer and actuation in the Smart Field Layer.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

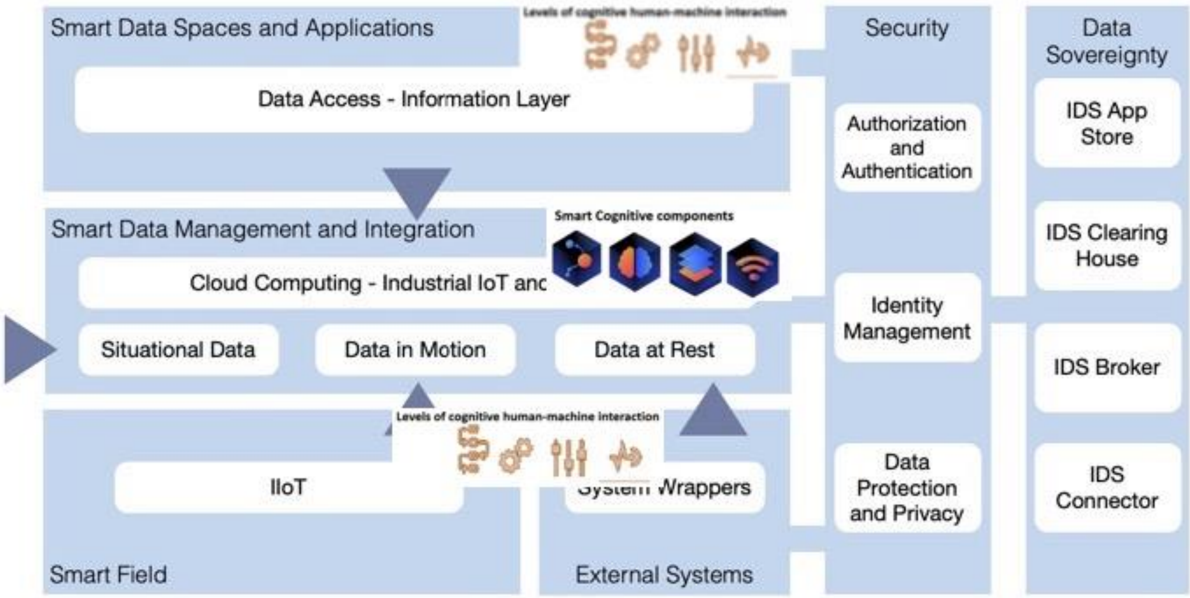

Figure 13 CAP Reference Architecture – Cognitive Mapping

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

4 CAPRI Open Source Reference Implementations

4.1 CAP Open-Source Implementation

The CAP Open-Source Implementation aims to provide a series of interoperable open-source tools (coming from FIWARE and Apache communities) that can be adopted the scope of lowering barriers during the development and deployment phases. Of course, these kinds of tools are not the only ones can be used, but they are considered as a starting point (easy onboarding) and any kind of integration could be done.

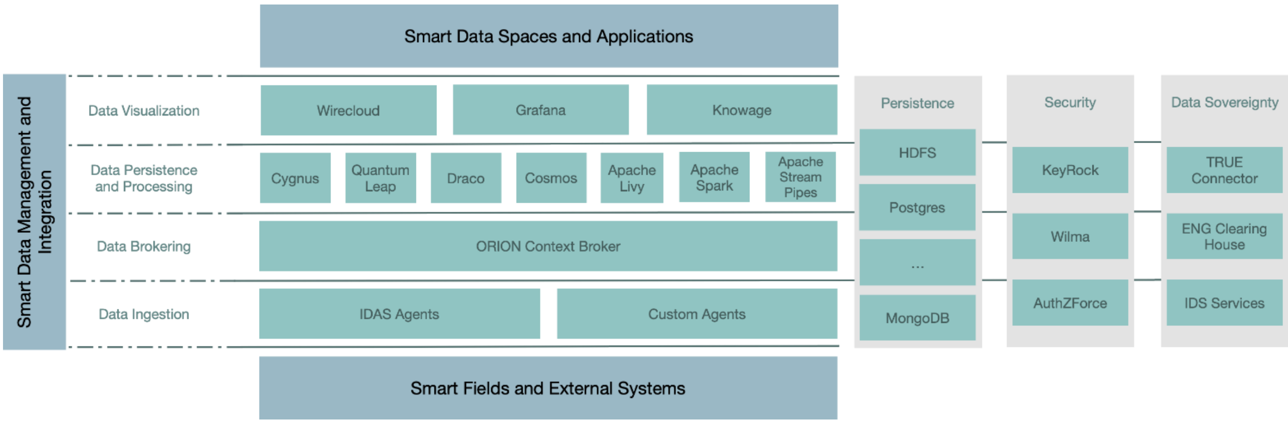

Figure 14 CAP Reference Implementation

Starting from the bottom there is the **Smart Field and External Systems** block, a physical layer where are located for instance the industrial devices, machines, actuators, sensors, wearable devices, robots, etc, such as the interfaces to collect data and communicating with IIoT systems. The External Systems is composed by all IT systems for supporting industrial processes. Custom interfaces and system wrappers are a crucial part of the component, aiming to share data using smart data models for representing information.

Just a level above, the Smart Data Management and Integration section is in charge of collect, process and visualize data, since this process is quite complex and needs to use many components, it is structure in 4 different sublayers: Data Ingestion, Data Brokering, Data Persistence and Processing, Data Visualization.

Data Ingestion

Data Ingestion provide a bridge between the physical layer and the data brokering, where the data from the devices are shared in a standardized structure with the broker, putting the information at the disposal of the tools will analyse them.

- **IDAS Agents:** The IoT Agent component allows to connect objects to gather data or interact with them, typical IoT use case scenario. It's needed in case of connecting IoT devices/gateways to FIWARE-based ecosystems. IoT Agents translate IoT-specific protocols into the NGSI context information protocol, that is the FIWARE standard data exchange model. IoT Agent for OPC UA, IoT Agent for JSON, IoT Agent for Ultralight are some IDAS Agent in FIWARE Catalogue.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

- **Custom Agents:** Any other custom Agent can be developed basing it on the same standard and features of the IDAS Agents.

Data Brokering

The Data Brokering sublayer is charge of feed the persistence and processing phase, where the main actor is the **ORION Context Broker**, an implementation of the Publish/Subscribe Broker Generic Enabler, able to manage the entire lifecycle of context information including updates, queries, registrations, and subscriptions. It based on NGSI-LD server implementation to manage context information and its availability. This GE allows to create context elements and manage them through updates and queries, and to subscribe to context information receiving a notification when a condition is satisfied, for example in case of context change.

Data Persistence and Processing

The core part is storing the data collected and process them. In the following, a list of main FIWARE and Apache open-source components.

- **Cygnus:** is a connector with the scope to persist context data sources into third-party databases and storage systems, creating a historical view of the context. It is based on Apache Flume, that is a data flow system structured on the concepts of flow-based programming. Built to automate the flow of data between systems, it supports powerful and scalable directed graphs of data routing.

- **Quantum Leap:** is a Generic Enabler focused on persisting historical context data into time-series databases such as CrateDB with reference to maintaining a scalable architecture and compatibility with visualization tools such as Grafana.

- **Draco**: is a connector used to persist on text data sources into other third-party databases and storage systems, creating a historical view of the context. Based on Apache NiFi, a popular framework for data management and processing from multiple sources, it connects the Orion Context Broker to a wide range of external systems such as MySQL, MongoDB etc. Another usage of Draco is filter and repost context data back into Orion.

- **Cosmos:** is a FIWARE Generic Enabler for big data analysis, it is composed by a set of tools (Orion-Flink Connector, Orion-Spark Connector, Apache Flink Processing Engine, Apache Spark Processing Engine, Streaming processing examples using Orion Context Broker) that help achieving the tasks of Streaming and Batch processing over context data.

- **Apache Livy:** is a service allowing easy interaction with a Spark cluster over a REST interface. Through it, can be easily submitted Spark jobs or snippets of Spark code, synchronous or asynchronous result retrieval, as well as Spark Context management, everything via a simple REST interface or an RPC client library. Apache Livy also simplifies the interaction between Spark and application servers.

- **Apache Spark**: is an open-source parallel processing framework for running largescale data analytics applications across clustered computers. It can handle both batch and real-time analytics and data processing workloads. It's part of a greater set of tools, including Apache Hadoop and other open-source resources for today's analytics community. In this way, it can be considered as a data analytics cluster computing tool. It can be used with the Hadoop Distributed File System (HDFS), which is a particular Hadoop component that facilitates complicated file handling.

- **Apache StreamPipes:** is a self-service (Industrial) IoT toolbox to enable non-technical users to connect, analyze, and explore IoT data streams. StreamPipes has an exchangeable runtime execution layer and executes pipelines using one of the provided wrappers, e.g. standalone or distributed in Apache Flink. Pipeline elements in StreamPipes can be installed at runtime - the built-in SDK allows to easily implement new pipeline elements according to your needs. Pipeline elements are standalone microservices that can run anywhere - centrally on your server, in a large-scale cluster or close at the edge.

Data Visualization

At the end of data storage and processing, the results of the analysis need to be visualized. The data visualization assumes a relevant role, because in this phase is fundamental have a clear idea of what the information means, giving it visual context through maps or graphs. This makes the data more natural for the human mind to comprehend. The mentioned platforms are powerful tools to do that in a simple way and at the meantime they are compliant with the most common data sources.

- **WireCloud**: offers a platform aimed at allowing end users without programming skills to easily create web applications and dashboards/cockpits. The purpose is to integrate heterogeneous data, application logic, and UI components (widgets) sourced from the Web to create new coherent and value-adding composite applications.

- **Grafana:** is a web application for analytics and interactive visualization. It provides charts, graphs, and alerts for the web when connected to supported data sources (MySQL, PostgreSQL, …). It is expandable through a plug-in system. End users can create complex monitoring dashboards using interactive query builders.

- **Knowage:** offers a complete set of tools for analytics, paying attention in particular at the data visualization for the most common data sources and big data. It has many modules (Big Data, Smart Intelligence, Enterprise Reporting, Location Intelligence, Performance Management, Predictive Analysis) to fit the needs of the consumers.

Smart Data Spaces and Applications

The final goal is the development of a smart data applications. The Smart Data Spaces and Applications, in fact, contains the system and user applications for presenting and consuming data. BI & Analytics, AR/VR, Chatbots & Virtual Assistants, Self-service Visualization and Generic Cognitive Applications are the main fields supporting and providing the requirements for developments.

On the right side of the Figure 14 there are three other blocks ensuring three different types of service to the CAP platform.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

Persistence

The persistence is a key component of a successful platform, it is needed to store all the relevant information data, making them accessible for the outside when it is required.

- **HDFS:** the Hadoop Distributed File System (HDFS) is a distributed file system designed to run on commodity hardware. It has many similarities with existing distributed file systems. However, the differences from other distributed file systems are significant. HDFS is highly fault-tolerant and is designed to be deployed on low-cost hardware. HDFS provides high throughput access to application data and is suitable for applications that have large data sets.

- **Postgres**: is a powerful, open-source object-relational database system. It has like main features the transactions with Atomicity, Consistency, Isolation, Durability (ACID) properties, automatically updatable views, materialized views, triggers, foreign keys, and stored procedures.

- **MongoDB:** is a distributed database, document-based, generic purpose for modern application and cloud. It stores documents in JSON format, supports matrix and nested objects, an advanced query language allows the user to filter data using whatever key in JSON document, having at the same time all the advantages of a relation DB like ACID transactions, the use of join in the queries and so on.

Security

In the security block are defined the components for the authorization and authentication of users and systems. They also integrate modules for data protection and privacy.

- **KeyRock:** is a FIWARE component for Identity Management. Using Keyrock (in conjunction with other security components such as PEP Proxy and Authzforce) it is added OAuth2-based authentication and authorization security to services and applications.

- **Wilma:** in combination with Identity Management and Authorization PDP GEs, adds authentication and authorization security to backend applications. Thus, only FIWARE users will be able to access Generic Enablers and other REST services. The PEP Proxy allows to programmatically manage specific permissions and policies to resources allowing different access levels to users.

- **AuthZForce:** is the reference implementation of the Authorization PDP Generic Enabler. Indeed, as mandated by the GE specification, this implementation provides an API to get authorization decisions based on authorization policies, and authorization requests from PEPs. The API follows the REST architecture style and complies with XACML v3.0.

Data Sovereignty

Data Sovereignty block contains the components of the IDS ecosystem able to exchange data in a secure way guaranteeing the technological usage control and the implementation of the data sovereignty principles.

- **TRUE Connector:** is one of the available open-source connectors based on IDS standards, it is a technical component to standardize data exchange between participants in the data space.

- **ENG Clearing House:** is an intermediary that provides clearing and settlement services for all financial and data exchange transactions.

- **IDS Services:** describe complementary services for deploying an IDS ecosystem. An example is the IDS Metadata Broker, an intermediary that stores and manages information about the data sources available in the data space or the IDS Identity Provider that offers a service to create, maintain, manage, monitor, and validate identity information of and for participants in the data space.

4.2 CAP Open Source in the Use cases

After the RA update, see chapter 2, and an iterative collaboration with the pilots, directly involved in the review, that validated the updates and shared their feedbacks starting for the as is use cased and the final desired outcomes, the next step is understand how to integrate the already existing models and algorithms. Typically, the algorithms are written in Matlab, it means that they are not readable from most of the tools described above, then in relation to the effort required, there are 3 options to proceed:

1. The one with the most effort is the translation from Matlab to Python or any other scripting language supported.
2. Another option is to export the Matlab code in a library that will be used at runtime, with the compatibility constraints with all OS.
3. The last and the one with less effort is to use the models and algorithms as a service, providing them the inputs and then collecting the outputs.

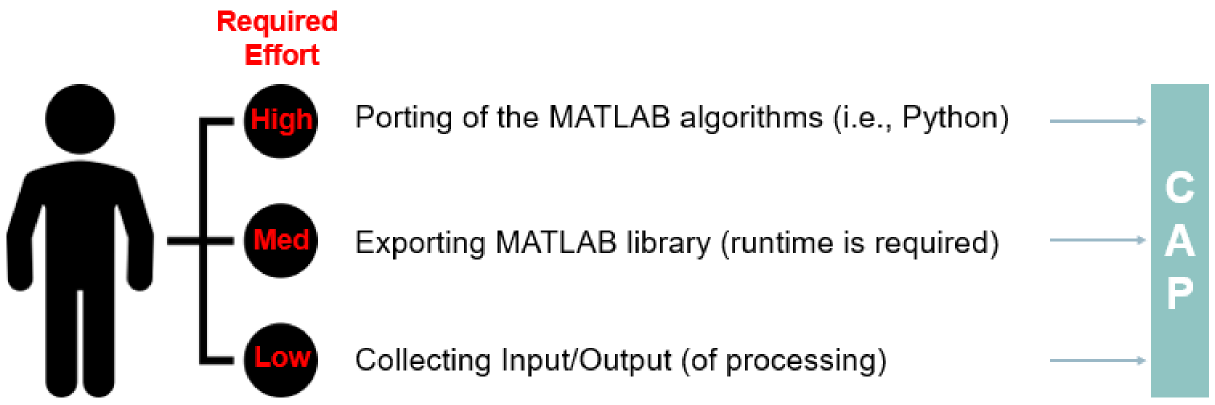

Figure 15 Model and Algorithm integration

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

The following sections define the mapping between each cognitive solution and the Open-Source implementation described above in chapter 2. For each cognitive solution and for each Case Study (Asphalt, Steel and Pharma), will be selected the most fittings components to deliver cognition capabilities in the scope of the specific solution and use case.

4.2.1 Asphalt use case

In relation to the Model and Algorithm integration already discussed in **¡Error! No se encuentra el origen de la referencia.**, in collaboration with the Asphalt pilot leader, **¡Error! No se encuentra el origen de la referencia.** is filled. The table reports several preliminary information about the technology used in case the code of the algorithm is subject to copyright and in case if it is feasible to convert in another programming language or export it as a library. Thanks to this contribution, for the algorithms already well defined, i.e., CAS1 or CAS2, it was taken the inputs to contaminate the related deployment of the CAP Reference Implementation. Following this path and constantly kept updated this approach for each algorithm, will be defined the next steps to better and better define the infrastructure to be used.

Table 1: Asphalt use case preliminary information

Algorithm	Description	Technology	Subject to copyright?	Can code be converted?	Can code be exported like a library (i.e., Matlab DLL)?	Must be run on the CAPRI PC (i.e., due to license issues)
CAS1	The code processes the spectral and image data to calculate the % of bitumen	C++ / CUDA	No	Yes	No	No
CAS2	Cognitive sensor of Filler Measurement	Labview, Phyton	No	No	Yes	Yes
CAC1	Cognitive control of the Asphalt Dryer Drum	MATLAB	No	No	As a Matlab executable by using the Application Compiler app and / or the Simulink Compiler for those machines that do not have MATLAB installed	Must be run in the system where asphalt plant data is being stored (according to our latest schematics, in the CAPRI server)
CAO1	Cognitive Predictive Maintenance	CORE to be detailed				

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

	of Baghouse Filter					
CAP1	Cognitive Planning and Optimization of Asphalt Production	R	No	No	Yes	Yes

In the Asphalt use case, data will be collected through a direct connection to WAGO data logger, using ad hoc customized connectors or, as an alternative/parallel option, using well known standards like MQTT. Once the data is collected will be persisted in Draco (a customization of Apache NiFi, already described in chapter 2) passing through Orion-LD, following the NGSI-LD standards. The data processing oversees Spark, enabled by Livy, that using its features, will provide an accurate dataset to Grafana to properly visualize the desired output. Data needs to be protected due to industrial property and for GDPR compliance, this part will be covered by KeyRock.

What is described above is a typical scenario thought to be deployed in order to satisfy the requirements related to the Asphalt use case.

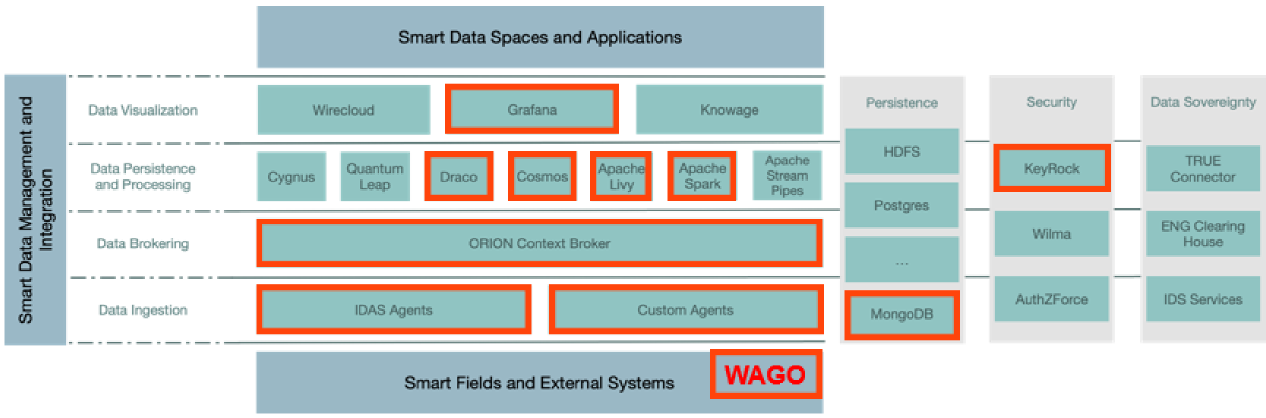

Figure 16 Asphalt use case open-source implementation

4.2.2 Steel use case

Also, for Steel use case is filled the **¡Error! No se encuentra el origen de la referencia.** in which is reported several information related to the technology used, if the code of the algorithm is subject to copyright and in case if it is feasible to convert in another programming language or export it as library. With respect to the previous use case, this time for most of the algorithms is feasible to export the code like a library, it doesn't mean that the integration can be easily done.

Table 2: Steel use case preliminary information

Algorithm	Description	Technology	Subject to copyright?	Can code be converted?	Can code be exported like a library (i.e., Matlab DLL)?	Must be run on the CAPRI PC (i.e., due to license issues)

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

CSS1	Optical tracking system for steel billets and bars	TODO	X	No	No	(Must be run elsewhere?)
CSS2	A simulation model for the continuous casting machine, calculates temperature and solidification front	C++, Python	X	No	Yes	No
CSS3	Tracks the temperature evolution of intermediate products in the rolling mill	Python	X	No	Yes	No
CSS4	Estimates the scale buildup during rolling	Python	X	No	Yes	No
CSS5	Estimates the processing risk for individual intermediate products (billets and bars)	Python (maybe some C)	X	No	Yes	No
CSO1	* Data access layer * Simulations * Visualizations & decision support	Python, NodeJS, Angular	X	No	Yes	No

In the Steel use case, the data collection is done by custom agent connecting from physical layer PLC, SCADA, MES systems to The Orion-LD Context Broker (following the NGSI-LD standards). When the data is collected the persistence is up to Draco (a customization of Apache NiFi, already described in chapter 2). Apache Livy will trigger the data processing invoking Spark's jobs, since not at all the algorithms can be used directly from Spark, if anyone of them is written in Matlab, an option is to feed that algorithm with data input e consume the output at run time.

At the end will be provided an accurate dataset to Grafana to properly visualize the desired output. Data needs to be protected due to industrial property and for GDPR compliance, this part will be covered by KeyRock.

What is described above is a typical scenario thought to be deployed to satisfy the requirements related to the Steel use case.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

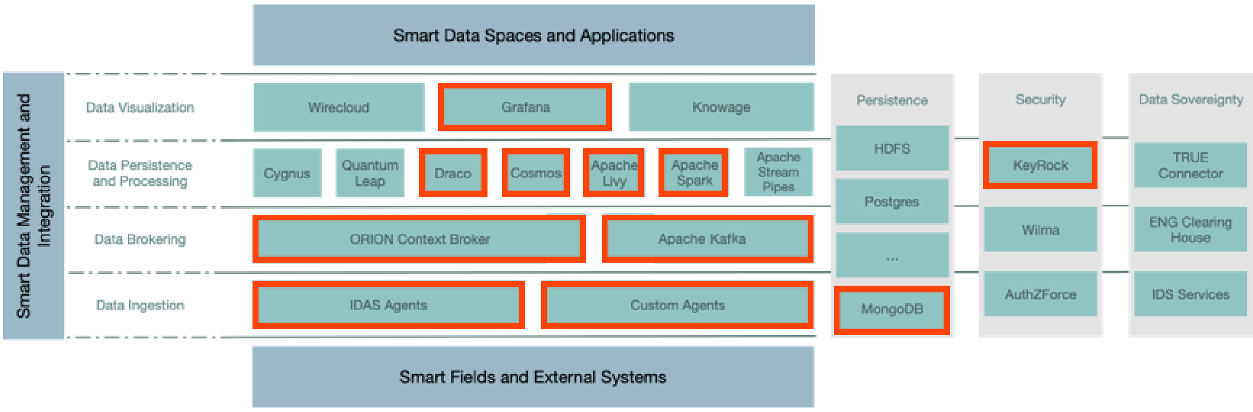

Figure 17 Steel use case open-source implementation

4.2.3 Pharma use case

An analogue work is ongoing also for Steel use case to classify in the **¡Error! No se encuentra el origen de la referencia.** the technology used, if the code of the algorithm is subject to copyright and in case if it is feasible to convert in another programming language or export it as library. As described in the table below, the technology mainly used in this case is Matlab, since in the most of the algorithms the conversion to another programming language cannot be done and either the exporting the code into a library is not allowed due to license reason, the idea to accomplish the service is to use a database where will be stored the inputs and after the processing the outputs through a REST API in order to allow Matlab to consume data on a specified machine autonomously.

Table 3: Pharma use case preliminary information

Algorithm	Description	Technology	Is code subject to copyright?	Can code be converted (i.e., target Java)?	Can code be exported like a library (i.e., Matlab DLL)?	Must be run on the CAPRI PC (i.e., due to license issues)
CPS1	Chemometric model for API content after granulation via Raman spectral data.	Matlab	Yes	Yes	Yes (we don't have the license for that)	No
CPS2	Determine robust characteristics of via Parsum acquired particle size distribution.	Matlab	Yes	No	Yes (we don't have the license for that)	Yes
CPS3	Development of an observer concept for granule moisture estimation in fluid bed dryer	Matlab	Yes	No	No	Yes
CPS4	Prediction of tablet dissolution profile	Not determined	Yes	Yes	Yes (we don't have	No

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

	from process and other CS data.	yet, likely Python			the license for that)	
CPS5	Algorithm for timely fault detection from process and other CS data	Not determined yet, to be discussed with Nissatech	Yes	Yes	Yes (we don't have the license for that)	No
CPC1	Model predictive controller for TSG and FBD, and model-based discharge strategy	Matlab	Yes	No	No	Yes
CPO1	Suggest actions to the operator, e.g., filter replacement, parameter adjustments	Not determined yet, to be discussed with Nissatech	Yes	Yes	Yes (we don't have the license for that)	No
CPP1	Automated DOE planning based on minimal energy and resource consumption	Not determined yet, likely Matlab	Yes	No	No	No

In the Pharma use case, the data will be collected through an IDSA Agent, for example the OPC-UA Agent, that is a plug and play component used to transfer and share contexts to Orion-LD from the OPC-UA servers at which it is connected. Whenever the data is collected, they will be stored in Draco (a customization of Apache NiFi, already described in chapter 2). The OPC-UA Agent, the OPC-UA servers and Orion-LD, following the NGSI-LD standards.

Apache Livy will start the data processing enabling Spark's jobs, then an accurate dataset will be provided to Grafana to properly visualize the desired output. Data needs to be protected due to industrial property and for GDPR compliance, this part will be covered by KeyRock.

What is described above is a typical scenario thought to be deployed in order to satisfy the requirements related to the Pharma use case.

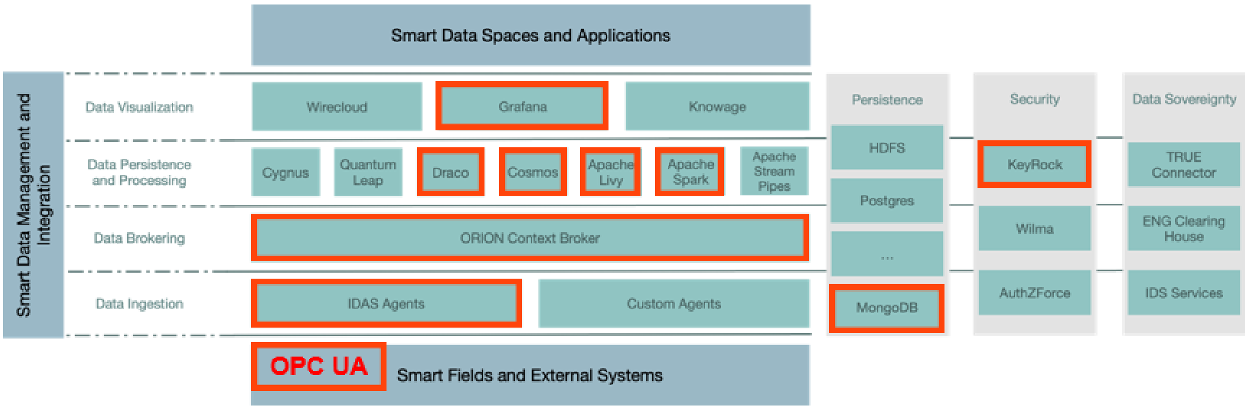

Figure 18 Pharma use case open-source implementation

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

5 CAPRI Data and knowledge reference implementation

This chapter aims at describing the reference implementation from the Data and Knowledge perspective. Data analysis, ontologies, Data4AI applicable to all CSs, with mapping from Reference Architecture to Implementation, based on data.

5.1 Pilots' dataset analysis

The goal is to present the works applied to pilots' dataset in the three CAPRI domains: Asphalt, Steel and Pharma.

5.1.1 Asphalt use case

We present an **initial** analysis in real data of ***predictive maintenance of baghouse (CAO1)***. More specifically, data from various sensors were collected and created an intermediate data set from the relevant features from April-5, 2021 to July-5, 2021. The obtained data concerned information for

- Baghouse filter drop pressure
- Drying drum drop pressure
- Baghouse temperature
- Electric Power blower
    - AMPERAGE Exhauster
    - ELECTRICITE Exhauster
- Information about production orders
- Maintenance history data

Notice that in the obtained data did not include any information about

- Exhaust power (%)
- Exhaust gases pipe drop pressure
- Dust emission in the clean gas chamber
- Combustion gases analyser

Initially the data were transformed and synchronized in order to develop a single intermediate dataset containing all sensor information, which is presented in Figure 19.

Date	Drying drum drop pressure	Baghouse filter drop pressure	Baghouse temperature	ELECTRICITE Exhausteur (KWH)	AMPERAGE Exhausteur (A)
2021-04-05 06:12:36	667	20131	71	4.4392	459.70830
2021-04-05 06:12:41	667	20398	70	4.4392	459.70830
2021-04-05 06:12:46	533	20531	70	4.4392	459.70830
2021-04-05 06:12:52	533	20531	70	4.4392	459.70830
2021-04-05 06:12:57	533	20398	70	4.4392	459.70830
...	...	...	...	...	...

Figure 19 Intermediate data set (CAO1)

Figures 20-24 present a line plot of the obtained numerical features i.e Baghouse filter drop pressure, drying drum drop pressure, Baghouse temperature, Electric Power blower, AMPERAGE Exhauster and ELECTRICITE Exhauster. At this point, it is worth mentioning that some temperature values were negative; thus, they were removed from our preliminary analysis. The interpretation of Figure 20 and Figure 21 suggests that the last 20 days (from June-15, 2021 to July-5, 2021) the values of Baghouse filter drop and Drying drum drop pressure were very close to zero. This is due to the fact that the system, which is registering pressure values, has change the measurement units. Thus, this change in units is responsible for presenting that both pressures seem to be zero compared with the pressure values from earlier days.

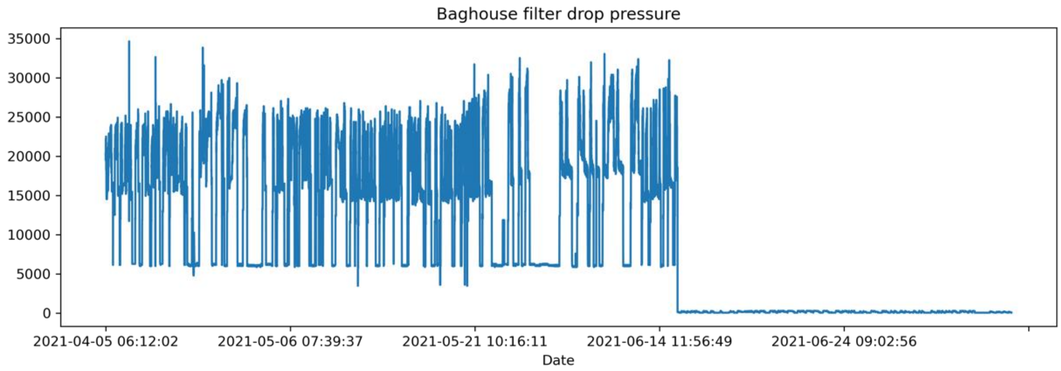

Figure 20 Line plot of Baghouse filter drop pressure

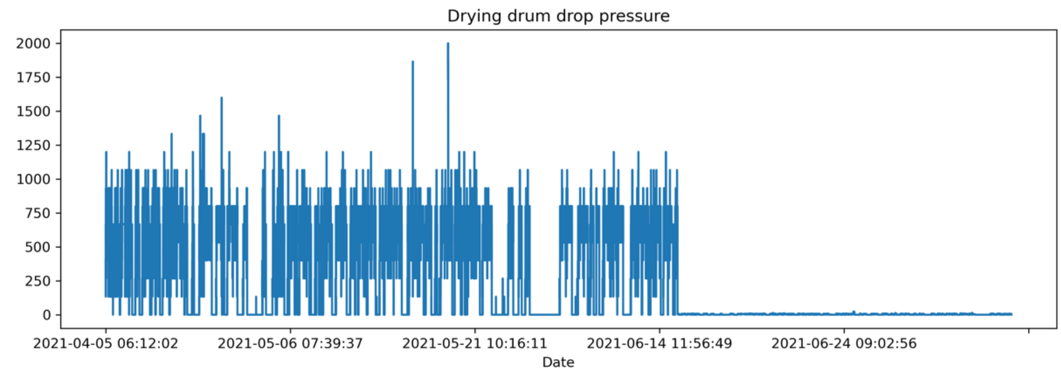

Figure 21 Line plot of Drying drum drop pressure

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

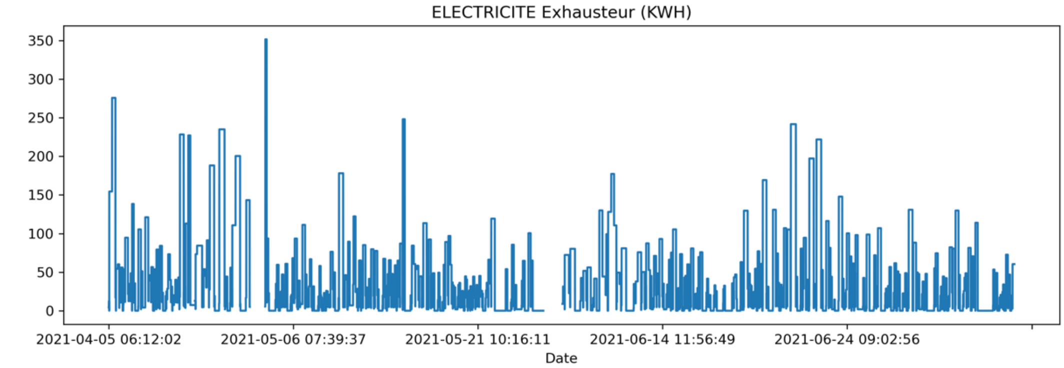

Figure 22 Line plot of ELECRICITE Exhauster (KWH)

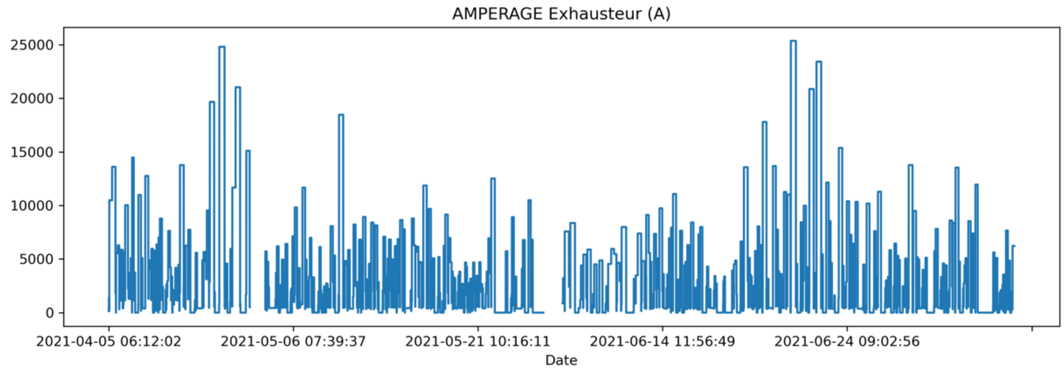

Figure 23 Line plot of AMPERAGE Exhauster (A)

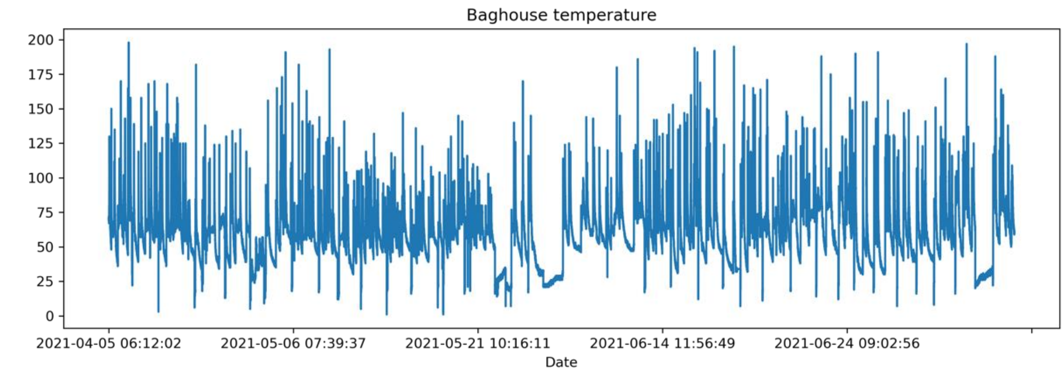

Figure 24 Line plot of Baghouse temperature

Finally, a correlation analysis was performed in order to identify and examine any linear dependencies between the relevant features. Figure 25 presents the correlation matrix of all pairs of features included in the data. Clearly, the features ‘drying rum pressure’ and ‘baghouse filter drop pressure’ are high correlated (0.90) as well as the features ‘AMPERAGE Exhauster’ and ‘ELECTRICITE Exhauster’ (0.91). It is worth noticing that any other pair of variables reported a correlation value up to 0.32, which suggests that any other pair of features presented low linear dependency.

This project receives funding in the European Commission’s Horizon 2020 Research Programme under Grant Agreement Number 870062

	Drying drum drop pressure	Baghouse filter drop pressure	Baghouse temperature	ELECTRICITE Exhausteur (KWH)	AMPERAGE Exhausteur (A)
Drying drum drop pressure	1.00	0.90	0.28	0.32	0.29
Baghouse filter drop pressure	0.90	1.00	0.17	0.26	0.21
Baghouse temperature	0.28	0.17	1.00	0.20	0.22
ELECTRICITE Exhausteur (KWH)	0.32	0.26	0.20	1.00	0.91
AMPERAGE Exhausteur (A)	0.29	0.21	0.22	0.91	1.00

Figure 25 Correlation matrix of features

The next steps, which will be performed in T3.3 are the following:

- Integration of information from production orders and historical data.
- Conduct a brief Exploratory Data Analysis with data distribution, boxplots, violin plots and also study the statistical properties of each obtained feature.
- Advanced anomaly detection techniques will be implemented in order to identify data points with abnormal behaviours.

5.1.2 Steel use case

5.1.2.1 Introduction

This report contains preliminary exploratory data analysis regarding **Sidenor** steel company, **Steel** use-case, **Capri** project. The main goal is to compare analysis performed by **D2Lab Online** software (performing Fast Thinking, see PICO architecture) and manually conducted analysis. It is crucial to understand that we are at this point still in the process of understanding nature of data itself and steps performed in production stages, so it is strongly advised to take this analysis only as a starting point.

5.1.2.2 Data description

Data consist of two stages:

- First stage consists of measurement data (fixed, non-volatile parameters, measured only once)
- Second stage consists of time series (parameters measured frequently at certain timestamps)

5.1.2.3 Analysis performed by D2Lab Online (Fast Thinking)

Analysis performed by **D2LabOnline** is displayed in Figure 26.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

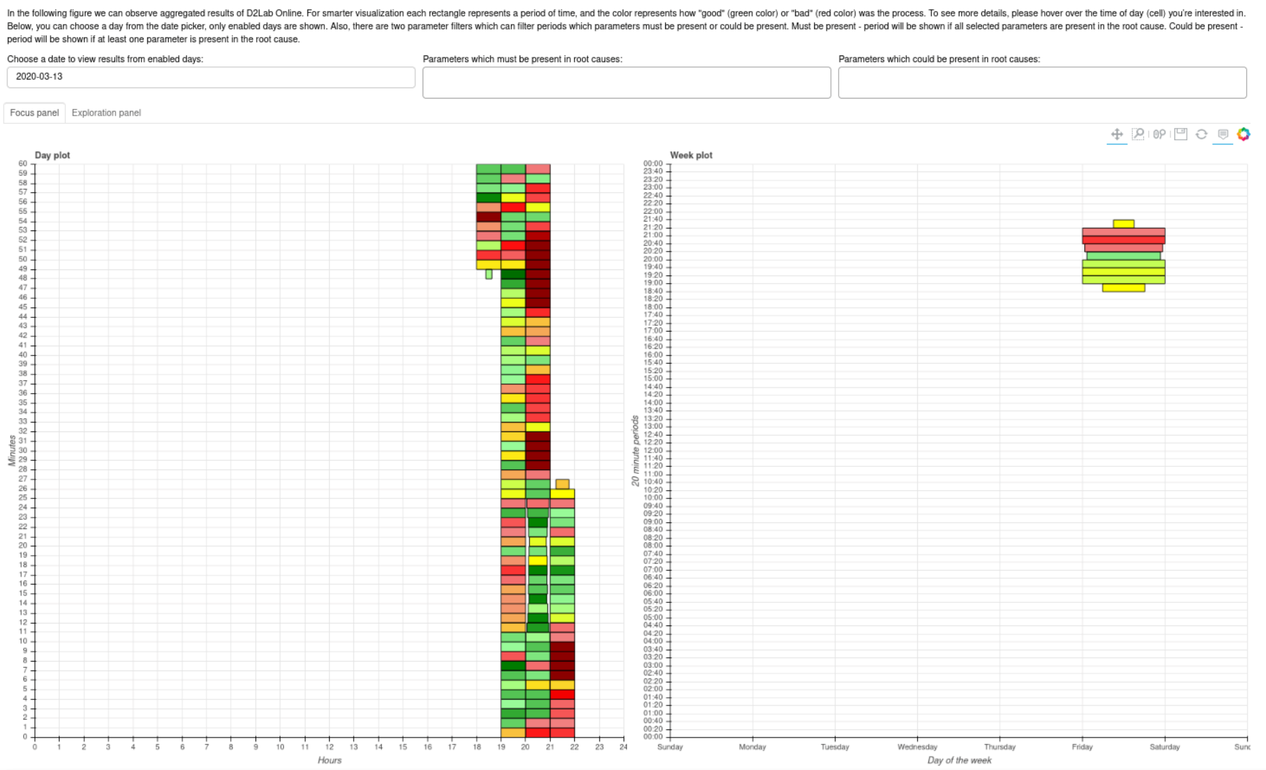

Figure 26 Analysis performed by D2LabOnline

By observing Figure 26, it is clear that only incomplete four hours of data are available, recorded on 13th March 2020. Dark green rectangles represent high satisfaction rate, opposite to dark red rectangles that represent low satisfaction rate (zero percent).

The main idea is to observe one "healthy" rectangle, represented with dark green color (satisfaction rate greater than 90 percent) as opposed to time interval containing sequential dark red rectangles (a few consecutive dark red rectangles).

The observed dark green rectangle represents a time interval of one minute, from 19:48 to 19:49 (image 2). Tooltip is highlighted by the blue rectangle (pops up when the green rectangle is hovered over). Satisfaction rate is high (96.67 percent).

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

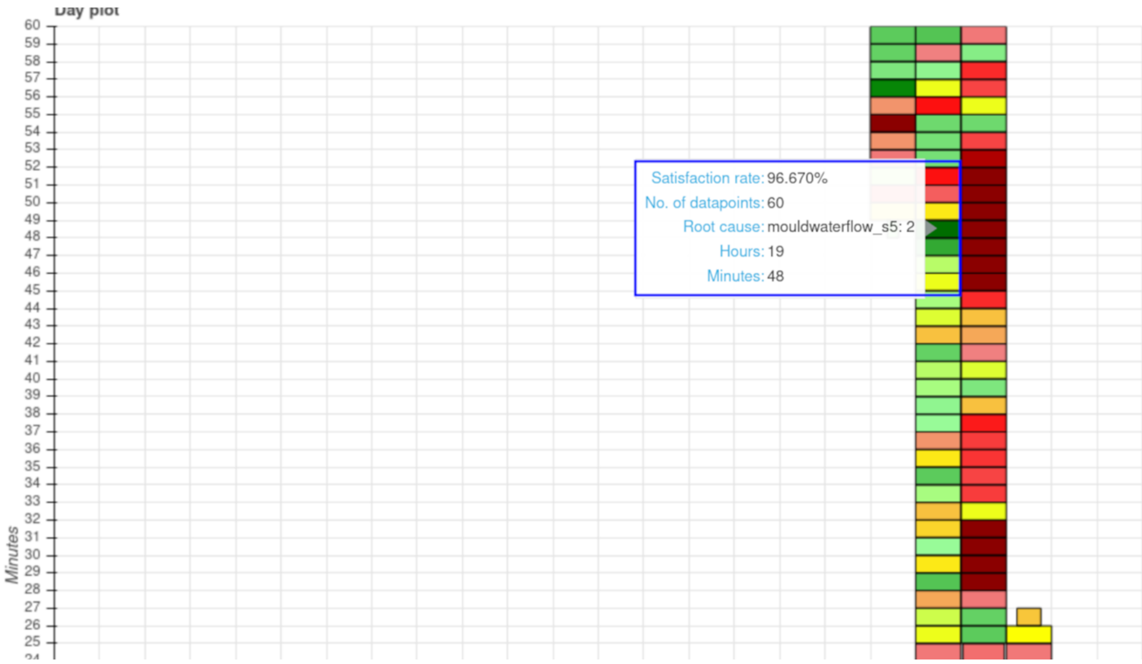

Figure 27 Tooltip representing time interval 19:48 – 19:49

As opposed to highlighted green rectangle represented in Figure 26, Figure 27 represents one dark red rectangle with satisfaction rate of 0 percent. This dark red rectangle also represents one minute interval 21:06 – 21:07.

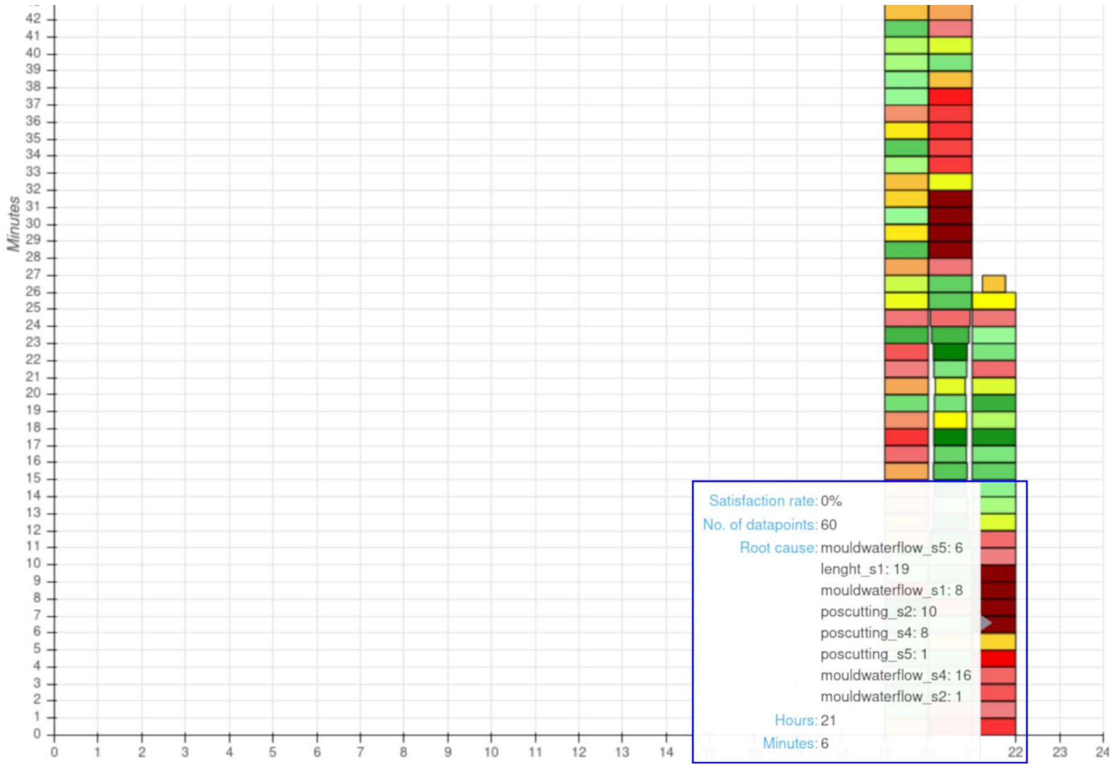

Figure 28 Tooltip representing time interval 21:06 – 21:07

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

5.1.2.4 Statistical parameters comparison

Figure 29 displays statistics for all parameters contained within both created frames. Upper frame displays statistics for parameters present in frame marked as four sequential dark red rectangles. Lower frame contains statistics for all parameters in frame marked with dark green rectangle. Due to the number of columns, only part of statistical parameters is represented in images.

```
df_red.describe()
```

	Unnamed: 0	instance_id	vacdegasduration	nsec	totsec	finStirringDuration	c	mn	si	s	p	ni	cr	mo	al	cu	sn	sb	as
count	241.000000	241.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
mean	2622.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	69.714896	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	2502.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	2562.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	2622.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	2682.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	2742.000000	210352.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
df_green.describe()
```

	Unnamed: 0	instance_id	vacdegasduration	nsec	totsec	finStirringDuration	c	mn	si	s	p	ni	cr	mo	al	cu	sn	sb	as
count	61.000000	61.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
mean	7322.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
std	17.752934	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
min	7292.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25%	7307.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
50%	7322.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
75%	7337.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
max	7352.000000	210351.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 29 Upper frame contains statistics for red rectangle, lower frame statistics for green rectangle

Main idea is to observe statistical parameters (mean, standard deviation etc.), derived from both created frames separately and compare statistics of corresponding parameters, as shown in Figure 30. For example, statistics of **steelTempMeniscus** parameter of red frame is highlighted by red rectangle, while statistics of same parameter in green frame is marked with green rectangle.

```
df_red.describe()
```

steelTempMeniscus	velCast_S1	lenght_S1	mouldWaterFlow_S1	mouldWaterTempDiff_S1	sprayWaterFlow_Z1_S1	sprayWaterFlow_Z2_S1	sprayWaterFlow_Z3_S
241.000000	241.000000	241.000000	241.000000	241.000000	241.000000	241.000000	241.0000
1521.352697	151.410788	1098.269253	2098.863071	6.841621	56.692946	32.435685	21.7925
14.111612	22.376171	1.779191	2.379147	0.169600	6.980114	2.819023	1.4224
1451.000000	111.000000	1094.990000	2093.000000	6.383101	45.000000	27.000000	19.0000
1525.000000	129.000000	1096.740000	2097.000000	6.759258	48.000000	29.000000	20.0000
1526.000000	154.000000	1098.460000	2099.000000	6.903934	59.000000	34.000000	23.0000
1526.000000	174.000000	1099.800000	2100.000000	6.932869	63.000000	35.000000	23.0000
1528.000000	178.000000	1101.070000	2104.000000	7.077547	64.000000	36.000000	23.0000

```
df_green.describe()
```

steelTempMeniscus	velCast_S1	lenght_S1	mouldWaterFlow_S1	mouldWaterTempDiff_S1	sprayWaterFlow_Z1_S1	sprayWaterFlow_Z2_S1	sprayWaterFlow_Z3_S1
61.000000	61.000000	61.000000	61.000000	61.000000	61.0	61.000000	61.0
1534.803279	169.524590	961.503230	2099.885246	7.198977	62.0	33.622951	22.0
0.400819	1.246416	0.502514	1.752126	0.036147	0.0	0.488669	0.0
1534.000000	167.000000	960.654000	2097.000000	7.135416	62.0	33.000000	22.0
1535.000000	169.000000	961.079000	2098.000000	7.193285	62.0	33.000000	22.0
1535.000000	170.000000	961.503000	2100.000000	7.193285	62.0	34.000000	22.0
1535.000000	170.000000	961.928000	2101.000000	7.193285	62.0	34.000000	22.0
1535.000000	172.000000	962.352000	2104.000000	7.251154	62.0	34.000000	22.0

Figure 30 Comparing statistics of corresponding parameters

5.1.3 Pharma use case

The general purpose of the Pharma dataset processing is to monitor the quality of pills processing. Main goal of the process is to perform the fault detection of the system and to alert the operator about the malfunction of the system for making the required corrections.

5.1.3.1 Received information

The received dataset contains total number of 20458 datapoints with sample rate of 1s, with 207 parameters measured in each datapoint. The total size of received dataset is 39.4 Mb. As the key index DATE column is defined.

The provided dataset contains data for approximately 6 hours, for time periods from 10.1.2020. at 5:38:37 until 10.1.2020. At 11:19:34. For specific time period, the following parameters does not contain any measurements, while all other parameters contain data for each datapoint:

Out of 207 provided parameters, 129 parameters are marked as important parameters for processing

5.1.3.2 Fast Thinking analysis

We present several examples of the analysis performed by **D2Lab Online** software (performing Fast Thinking, see PICO architecture).

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

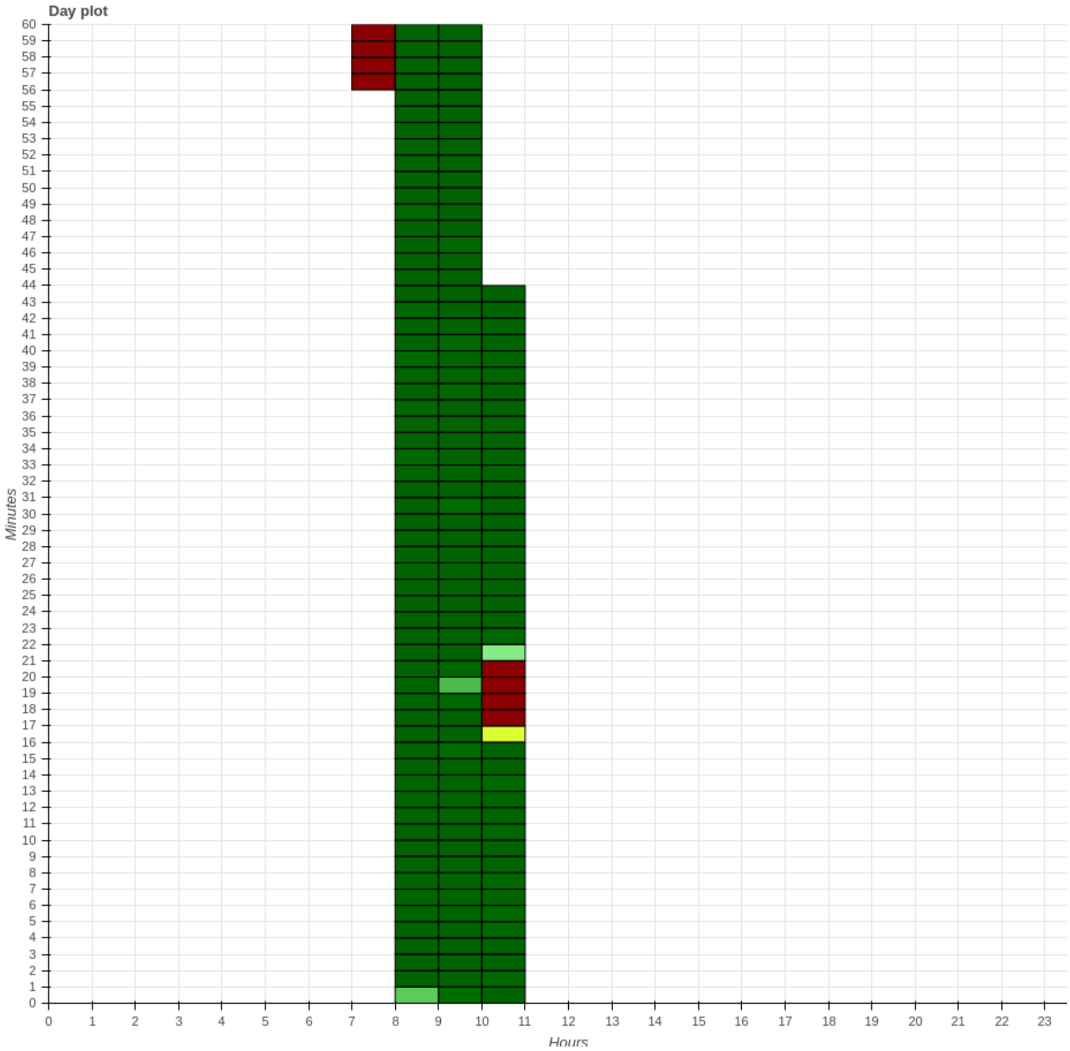

Figure 31 D2Lab Online report for Dryer

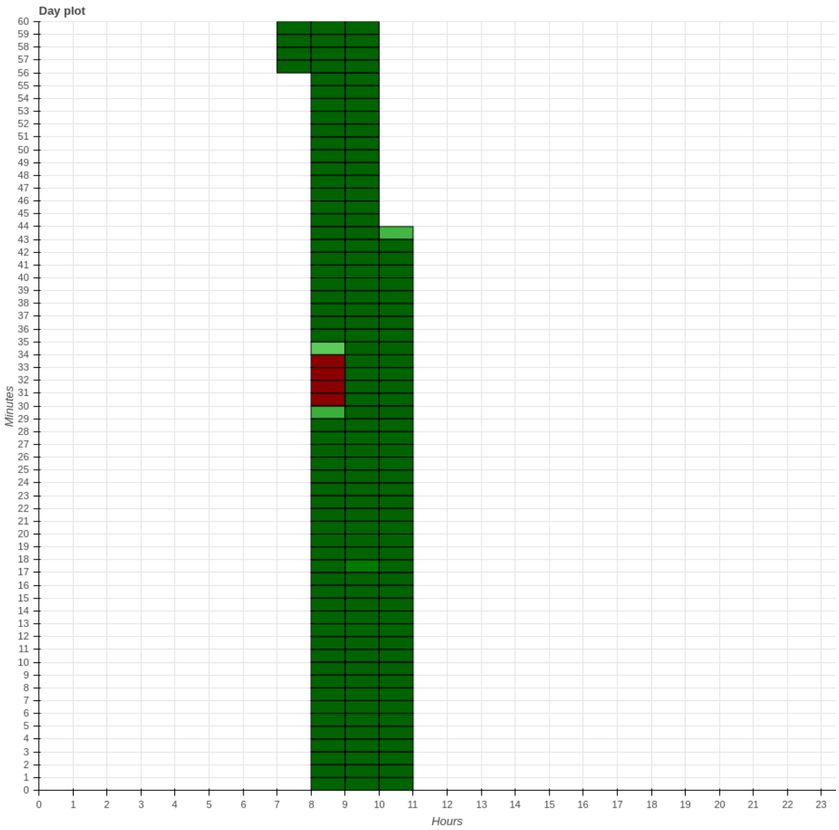

Figure 32 D2Lab Online report for Tablet press

5.2 Ontology

In this section we provide a short analysis of the ontologies used in pilot domains.

The analysis done by pilot partners has shown that only Pharma domain has experienced the usage of ontologies. The area is Risk management.

Risk management is a key element of the quality by design (QdB) approach in pharmaceutical manufacturing [doi: 10.1208/s12248-014-9598-3]. In [doi: 10.1016/j.eswa.2012.01.089], a risk management ontology is proposed: A development strategy is presented that covers specification, conceptualization, formalization and implementation. The specification covers the risk management tools to be used, the development approach to be implemented, the intended users and target groups of the ontology, the technical specifications (e.g., the implementation framework) and the functional requirements of the ontology. One of the tools commonly used is the Ishikawa diagram. For the process considered in the present project (ConsiGma 25 continuous from powder to tablet line), an Ishikawa diagram has been created and is shown in Figure 33.

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

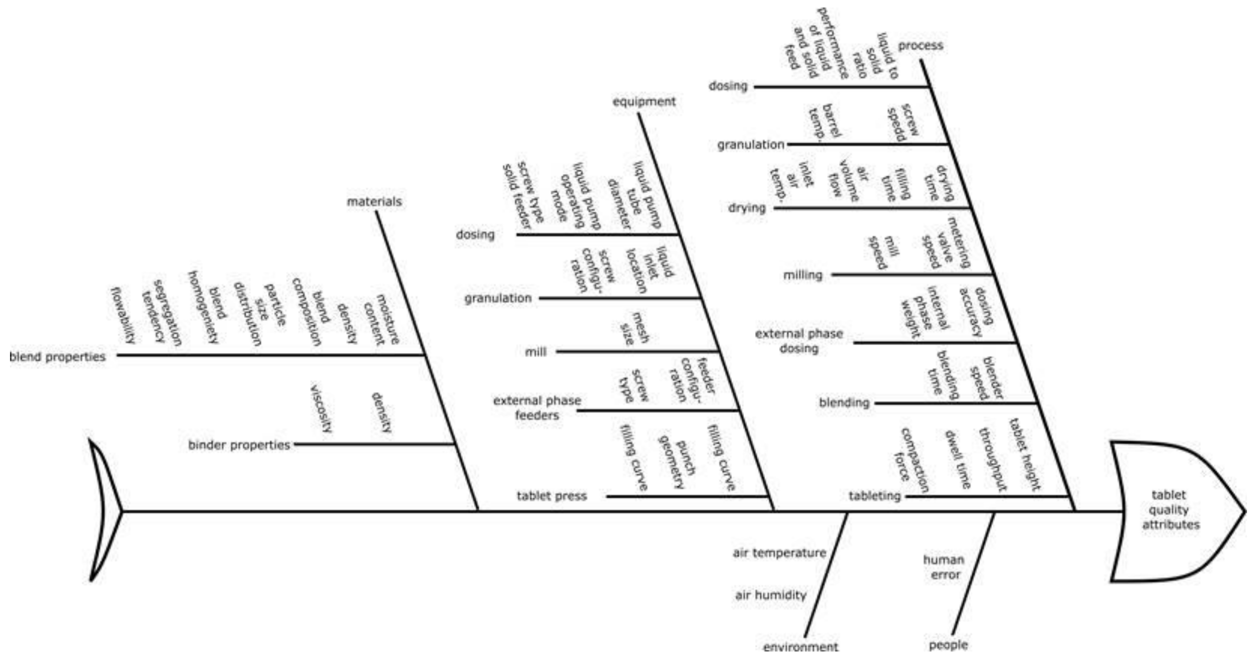

Figure 33 Ishikawa diagram of the considered process

The conceptualization phase then deals with the definition of a common vocabulary, as well as with the definition of all main structural components and their constraints. The ontology formalization phase then explicitly describes all structural components and their constraints, e.g., via an object-oriented technology like UML (Unified Modeling Language). In the implementation phase the ontology is then encoded by means of an ontology editor (e.g., Protégé), an ontology markup language (e.g., OWL DL) and an ontology query language (e.g., SPARQL). Finally, a verification and validation of the ontology has to be performed. More details on the above mentioned procedure can be found in [doi: 10.1016/j.eswa.2012.01.089].

This project receives funding in the European Commission's Horizon 2020 Research Programme under Grant Agreement Number 870062

6 CONCLUSION

Deliverable D3.1 – “CAPRI final reference architecture” is expected to be an extremely important document for CAPRI project since it collects the requirements for the CAP implementation in light of what has been achieved in a year and an half of laboratory activities in WP3.

These requirements have been used as starting point to sketch the high-level description of the main components to be implemented in the final solution.

Actually, D3.1 represents the bridge that connects WP3 – “Smart modules for cognitive process industry plants” and WP4 – “Cognitive technology solutions for process industry plants”, since the features of the Cognitive Automation Platform that have been identified in D3.1 will be then implemented in WP4.

The definition of the Reference Implementations for each use case has paved the way for the elicitation of common requirements that the CAP platform will satisfy in the context of the WP4.
In particular, cognitive solutions (i.e., sensors, control, operation, and planning) will be implemented and integrated in order to satisfy the needs of the three CAPRI domains (i.e., Asphalt, Steel, and Pharma) supporting all use cases and covering the entire life data cycle from the data ingestion to the data presentation. The algorithms, already analysed in terms of rationale, technology, and intellectual property, will be integrated into the CAP platform for implementing the processing layer. After the integration of the cognitive modules, the platform will be tested and tuned, making it ready for the validation scenarios (WP5) through two iterations (the first one is planned for M24).

WP3 will start a strong collaboration not only with WP4 for the implementation and WP5 for the validation, but also with WP7 in order to define the unique value proposition of the CAP and its original aspects that make the solution innovative.

In addition, D3.1 deal with other two main topics: the PICO (Process Industry Cognition) Architecture and Pilots’ dataset analysis run so far, still at an initial stage but planned to be improved and enhanced. The former has been presented in form of high-level description of the architecture with great chance of applicability in CAPRI context, since it is designed to perfectly match with CAPRI Reference Architecture, encompassing the four Smart cognitive components. The latter deals with dataset analysis that so far has been performed separately for each domain: next step is to generalise the analysis, finding out common features in the three domains.

This project receives funding in the European Commission’s Horizon 2020 Research Programme under Grant Agreement Number 870062