capri

Cognitive Automation Platform for European PRocess Industry digital transformation

Deliverable

D3.4 CAPRI Smart knowledge and semantic data models

Deliverable Lead: CARTIF

Deliverable due date: 31/03/2022

Actual submission date: 31/03/2022

Version: V6

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

Document Control Page	
Title	CAPRI Smart knowledge and semantic data models
Lead Beneficiary	CARTIF
Description	This deliverable of type OTHER presents the implementation results of the data-driven Reference Architecture for Cognitive Plants in the process industry, with smart knowledge and semantic data models cognitive-based components, as developed in T3.3. The Capri Reference Implementation, as described in D3.1, has been materialized by Open Source environments such as APACHE and FIWARE, and state of the art open standards such as OPC UA and MQTT. Final outcomes will be provided as open from CAPRI ZENODO account, the present report for each Cognitive solution includes the following details: • **Text document** (Word/Pdf): explaining the work done, introduction to video, user manual of the apps, including the available links, data and metadata description, description of everything else in the ZIP. • **Executable of the program/app**, if any • Any kind of **data** that can be publicly shown (full data, sample data, metadata, surveys results, …) • **Video** used to explain “physical” results, as sensors.
Contributors	Cristina Vega (CAR), Mireya de Diego (CAR), Laura Sanz (CAR), Aníbal Reñones (CAR), Antonio Salis (ENG), Silvia Razzeti (POLIMI), Jakob Rehrl (RCPE), Spyros Fourakis (CORE), Christoph Nolle (BFI)
Creation date	17/01/2022
Type	OTHER
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

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

Document History			
Version	Date	Author(s)/ Reviewer(s)	Status
0.1	17/01/2022	Cristina Vega (CAR)	First ToC
1.0	25/02/2022	Cristina Vega (CAR), Mireya de Diego (CAR), Laura Sanz (CAR)	Ch1, 3.1 contributions
1.1	28/02/2022	Silvia Razzeti (POLIMI)	Ch 2 contribution
1.2	02/03/2022	Spyros Fourakis (CORE)	CORE contribution to CAO1
1.3	06/03/2022	Christoph Nolle (BFI)	BFI contribution
1.4	11/03/2022	Jakob Rehrl (RCPE)	Pharma Contributions
2.0	11/03/2022	Cristina Vega (CAR), Aníbal Reñones (CAR)	Review contents
3.0	14/03/2022	Antonio Salis (ENG)	Review contents
3.1	17/03/2022	Spyros Fourakis (CORE)	CORE contribution review
4.0	21/03/2022	Jakob Rehrl (RCPE)	Pharma Contributions review
5.0	29/03/2022	Cristina Vega (CAR)	Zenodo links, last updates
6.0	31/03/2022	Cristina Vega (CAR)	Overall review

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

Table of Contents

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

This report is focused in the explanation of the different assets shared as OTHER type of deliverable. It contains, data, algorithms, videos, models etc. shared that represents the current status of the cognitive solutions of the CAPRI project that fall under the category of “Smart knowledge and semantic data models”. The status reflects the works done at laboratory level previously to be integrated into actual manufacturing environment of industrial use cases participating in the project.

CAPRI wants to develop cognitive solutions to help process industries. The cognitive solutions shown here belong to the ones used at **Operational level**. These solutions intend to help detect changes in the production process so the plant is able to respond to these dynamic fluctuations by adapting the production to stay within the targets of costs and rate, as well as quality and sustainability. Some examples of Operational level cognitive solutions the reader will see in this report include a Predictive maintenance cognitive solution for Asphalt baghouse (CAO1), a digital twin for the production of steel bars (CS01) and a cognitive operation concept for the production of pharma tablets (CP01).

The reader has access to the different assets generated included inside a ZIP file but also contained at CAPRI account of Zenodo open science repository and CAPRI’s YouTube channel for certain results which are protected but which functionalities can be appreciated through a series of instructive videos.

1 Introduction

1.1 Scope of Deliverable

The deliverable D3.4 CAPRI Smart knowledge and semantic data models, deals with explaining the work performed during the development of task T3.3 of the CAPRI Project. The Process industry needs to be modelled and simulated, and in this deliverable, it will be explained per use case, and per cognitive solution, the different models and the used simulation tools, like digital twins, for the development of those operational cognitive solutions that, focusing on the products, machinery or processes, achieve an enhanced and optimized operation with expert knowledge from workers and real time information, usually at plant level. In case of using an ontology, its explanation will also be included here.

Each cognitive solution of operation will be described with the same structure, starting with a general explanation of the developed solution to a deeper, more detailed description of the model, algorithm and/or the generated data.

1.2 Audience

Target audience of this deliverable are those industry sectors that wanted to implement and commissioning different cognitive solutions where, at an operational level, wanted to achieve these different goals:

- Changes in the production process will be detected and the plant will be able to respond to these dynamic fluctuations by adapting the production to stay within the targets of costs and rate, as well as quality and sustainability thresholds.
- High-level business intelligence for the whole plant, focusing on the economic peculiarities.
- Decision support tools to reroute or cancel processing of those faulty produced materials or products or reallocating these products to alternative orders.

This way, all those interested in these issues will find in this deliverable several operational cognitive solutions where its development methodology can be used as a basis for solutions adapted to their particular cases.

1.3 Relationship with other deliverables

The scope of this deliverable is focused on those Cognitive Solutions (CS) developed within the CAPRI project focused only on the Smart knowledge and semantic data models. Due to that it is complicated to separate and classify each one of the developed CS in only one kind of solution, i.e.: sensors, control, operation or planning; each CS will be part of one type of this classification. but taking into account that many of the CS will be part with more or less proportion in some of the others points of the mentioned classification. In this deliverable, the point of the classification of the CS is that one that is part of the smart knowledge and semantic data models, which are those ones related to operational solution. There is no single developed CS that can be classified at only one level.

Here, in the chapter two, those CS which have been classified at a large percentage to be within the operational level are described in this deliverable.

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

1.4 Document Structure

This document is structured in 5 chapters:

First one is the introduction chapter where the contents of this deliverable are introduced.

After that, the technical part starts, divided in 3 chapters, to end with chapter 5 focused on the conclusions.

Chapter 2 gives an overview of the structure of the deliverables of type OTHER that are part of this deliverable.

Following this, chapter 3 main objective is the explanation of the smart knowledge and semantics data models developed in each use case, analysing all the involved CS (cognitive solutions) from a general perspective.

Next, chapter 4 gives a detailed description of each one of the CS that have been developed in CAPRI project at operational level. There are 4 levels of the cognitive solutions, from sensors to planning, passing for control and operation too. This chapter describes all those ones aimed at an operational level, giving a high-level description of the model / algorithm of the corresponding CS and a description of the different files as part of this deliverable.

2 Attachment Structure Description

The objective of the current chapter is twofold: first, it aims to provide an overview of the structure of the deliverables of type OTHER, including the current one; secondly to provide the list of the documents attached to the present report, including the Zip folder, the Zenodo folder and any other external link (e.g. CAPRI youtube channel for the case of videos).

Actually, at month M24 (March 2022), WP3 is delivering four deliverables of type OTHER to summarise and provide evidence of what has been implemented so far; each deliverable consists in a textual part (the present document) and a set of attachments, which is the core of the deliverable.

This section aims at providing a summary picture of the structure of the four documents and of the attachments to make easier to find relevant information, since the subjects presented in each report may overlaps with the others and it is not straightforward for the reader to understand what he/she can find where. Hence, the objective of this chapter is to provide support in orientating inside them.

- Deliverable D3.2 – “CAPRI Industrial IoT Platform and Data Space” as output of Task 3.1;
- Deliverable D3.3 – “CAPRI Industrial Analytics Platform and Data Space” as output of Task 3.2;
- Deliverable D3.4 – “CAPRI Smart Knowledge and Semantic Data Models” as output of Task 3.3, the current one;
- Deliverable D3.5 – “CAPRI Smart Decision Support” as output of Task 3.4.

As mentioned above, each deliverable is related to a specific Task, but the activities performed in WP3 cannot be isolated per Task since they involve more than one at the same time. Actually, the CAPRI Cognitive Solutions are 19 assets implemented at laboratory level, split by the three project domains (Asphalt, Steel and Pharma) and encompassing the four layers of WP3 (Sensor, Control, Operation and Planning, corresponding to the four Tasks). It means that each CS is developed within one specific use case, but it presents features that cross more than a layer, so practically, it is part of more than one Task.

To overcome this situation and the fact that each CS should be described in more than one deliverable, it has been agreed to include in the report an initial section (Chapter 3) to describe the activities associated to the Task. So, Chapter 3 is at Task level and takes care only of the CS’s component related to the task, even if it means to depict it only partially.

Conversely, Chapter 4, that is the core of the deliverable together with the set of attachment, provides the overview at CS level, in order to avoid jumping from a document to another to find information. Each Chapter 4 (of the four deliverables) contains only a subset of CSs but they are fully described: for each cognitive solution, the main achievements are presented.

The following table shows in which way the 19 Cognitive Solutions encompass the four WP3 layers and so, also the four WP3 Tasks (in bold, the percentage that drove the choice of the deliverable where the CS has been assigned).

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

Table 1 CAPRI CSs encompassing the four layers/tasks

DOMAIN	CS'S NAME	CS'S CODE	T3.1 Sensor	T3.2 Control	T3.3 Operation	T3.4 Planning
ASPHALT	Sensor for bitumen content	CAS1	**50%**	50%		
	Sensor sensor of amount of filler	CAS2	**80%**	10%		10%
	Control of the asphalt drum	CAC1		**85%**	15%	
	Predictive Maintenance of baghouse	CAO1	10%		**30%**	60%
	Planning and control of asphalt production	CAP1		10%	10%	**80%**
STEEL	Sensor for product tracking	CSS1	**70%**	30%		
	Sensor for Solidification	CSS2	20%	**80%**		
	Sensor for Product temperature	CSS3	20%	**80%**		
	Scale sensor for scale build-up	CSS4	20%	**80%**		
	Sensor for risk and anomalies	CSS5			30%	**70%**
	Digital twin architecture	CSO1		10%	**60%**	30%
PHARMA	Sensor for blend uniformity	CPS1	**80%**	20%		
	Sensor for granule quality	CPS2	**80%**	20%		
	Sensor for product moisture	CPS3	10%	30%	**60%**	
	Sensor for prediction of dissolution	CPS4		40%	**60%**	
	Sensor for fault detection	CPS5		10%	**60%**	30%
	Cognitive Control Concept	CPC1		**70%**	30%	
	Cognitive Operation Concept	CPO1		10%	**70%**	20%
	Cognitive Planning Concept	CPP1		10%	10%	**80%**

All project partners have agreed to assign each CS to a specific deliverable, even if it encompasses more than a layer (and so, more than a Task) to ease and speed-up the reading of the document and to show all the information related to a Cognitive Solution in a single report.

Hence, the 19 CSs have been split in the four deliverables as follow, according to the most relevant layer:

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

Figure 1 The 19 CSs into the deliverables D3.2, D3.3, D3.4, D3.5

In this way, the four deliverables are well balanced: 5 CSs are described in D3.2 and D3.3, whose main component is the Sensor and the Control, respectively; 6 CSs are described in D3.4, focused on Operation and finally, 3 CSs are described in D3.5, about Planning. In each deliverable, Asphalt, Steel and Pharma domains are always represented.

Namely, the current deliverable contains the following 6 Cognitive Solutions:

- CAO1 – Predictive Maintenance of baghouse [Asphalt]
- CSO1 – Digital twin architecture [Steel]
- CPO1 – Cognitive Operation Concept [Pharma]
- CPS3 – Sensor for product moisture [Pharma]
- CPS4 – Sensor for prediction of dissolution [Pharma]
- CPS5 – Sensor for fault detection [Pharma]

Since we are talking of deliverables of type OTHER, each Cognitive Solution listed above is equipped with:

- A number of attachments of different nature (video, data, metadata, …), containing additional information that helps to better understand the final output of the CS and to provide a concrete evidence of what has been implemented in WP3;
- A textual part, available in Chapter 4, which complements the "physical" content in attachment, explaining what it is and how to exploit it.

The attachments are available in the Zip Folder, in Zenodo folder or as a YouTube video, according to its format. Due to limitations of space (52Mb) in EC portal not all the assets could be included into 1 single file. That is the reason we have decided to include all files into CAPRI's Zenodo account and CAPRI YouTube channel video for the videos showing specific demonstrations. The table below lists all the links (zenodo and youtube) of the different files described in the present report.

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

Table 2 D3.4 – List of attachments

Cognitive Solution	Content	Type	Location
CAO1 – Predictive Maintenance of baghouse	CAO1_Image_1.png CAO1_Image_2.png CAO1_Image_3.png	Picture	https://doi.org/10.5281/zenodo.6393690
CSO1 – Digital twin architecture	CSO1_API_1.pdf	Pdf	https://doi.org/10.5281/zenodo.6393030
	CSO1_video_1.mp4	Video	https://doi.org/10.5281/zenodo.6393030
CPO1 – Cognitive Operation Concept	CPO1_App_and_data_1.zip	Python code + Sample data	https://doi.org/10.5281/zenodo.6393041
CPS3 – Sensor for product moisture	CPS3_data1.zip	Process data + CPS3 output	https://doi.org/10.5281/zenodo.6393707
CPS4 – Sensor for prediction of dissolution	CPS4_App_and_Data_1.zip	Python code of CPS4, example dataset for CPS4	https://doi.org/10.5281/zenodo.6393975
CPS5 – Sensor for fault detection	CPS5_Data_1.zip	Sample dataset	https://doi.org/10.5281/zenodo.6394080

It is worth to mention that each data file shared is accompanied by the corresponding data management plan file (DMP file) to comply with the F.A.I.R.[1] principles as CAPRI project is part of the open research data pilot of H2020[2].

Finally, an overview of each Cognitive Solution is available in the CAPRI website, at the “Use Cases” section[3].

If you are interested in the details other CSs different from the six associated to the current deliverable, please refer to another deliverable according to the structure shown in Figure 1.

[1] Under these principles, each data file must be Findable, Accessible, Interoperable and Reusable. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

[2] https://www.openaire.eu/what-is-the-open-research-data-pilot

[3] https://www.capri-project.com/technology

3 Smart knowledge and semantic data models

3.1 Asphalt smart knowledge and semantic data models solutions

With the main objective of providing asphalt process use case with increased performance across different indicators and quality control of intermediate flows, the cognitive operational solution developed within the different cognitive solutions based on smart knowledge and data model platform for the asphalt use case is the following:

Predictive maintenance of baghouse based on cognitive sensors and expert knowledge.

A predictive maintenance system based on different process measurements like drop pressure between the drum and the filter in connection with the cognitive solution of the control of the rotary dryer drum is able to reduce maintenance costs and spare costs (filter bags mainly). The climatic conditions in the baghouse can produce a premature wear of the filters. When humidity is high, filler can make a coating layer with it and then, decrease baghouse filter performance and the durability of the corresponding filter bags and if a proper maintenance plan is not performed, even high temperature can happen making even the filters burn. The integration of historical maintenance intervention and maintenance operator’s knowledge data enhance the robustness and reliability of this predictive maintenance system.

Baghouse filter is a dust collector that removes dust, mainly filler content in dry aggregates during the corresponding drying process in the rotary drum. Baghouse performance is dependent upon inlet and outlet gas temperature and flow speed, pressure drop, opacity climatic conditions in the bag filters (temperature and humidity, asphalt mix recipe…). Performance degrades over time due to dust accumulation, so a preventive maintenance (periodic cleaning or changes) approach is needed. Power consumption varies over time in case bag filters get dirtier. Other reliability issues include pressure sensor faults, dust emissions due to broken or wear filter bags and so on. When the pressure between input and output varies a lot it can be an indicative that some filters might be deteriorated.

Based on big data analytics, historical maintenance data (which has the inconvenience that is mainly written down by hand), with a proper data merging coming from the main sensors related to the baghouse filter and its corresponding cleaning and pre-processing (removing outliers, means, etc.), an anomaly detection model has been developed. From this solution, an exploratory data analysis is performed and, depending on the calculated thresholds, several warnings or alarms can be raised to warn the operator(s)/manager(s) to perform proper maintenance operations. The proper knowledge-based algorithms have been developed based on machine learning algorithms based on an auto-encoder algorithm approach. An auto encoder is a type of artificial neural network used to learn efficient coding of unlabelled data (unsupervised learning). Then, the classification is validated and refined by attempting to regenerate the input from the encoding.

The approach from the knowledge base solution in the case of the asphalt plant use case is to identify possible malfunctions in the operation of baghouse and an estimation of the remaining useful life of a component of baghouse. More specifically, the proposed solution is composed by two models fed with the relevant process variables, provide a graphical representation of the evolution of the residual useful life of the component of baghouse and warning or alerts if the machine is not working properly.

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

3.2 Steel smart knowledge and semantic data models' solutions

In the CAPRI steel use case the concept of Digital Twins is used to provide a product-centric view on the process data. The starting point for the Digital Twin (DT) is the so called „digital shadow "(DS). This is a kind of a digital footprint of a physical object, in our use case a product or a production plant. The DS is a model which is fed by a one-way data flow with the state of the object. A change in state of the object leads to a change in the digital object, but not vice versa. The DS therefore transfers a physical object into the virtual world, leading to a sufficiently accurate digital image of the object, as illustrated in Figure 2.

Figure 2: A digital twin of a steel long product

Since the sensor data from the automation system is usually time-related, a conversion or an interpretation of the data needs to be done by the twin, either upon ingestion or on the fly at retrieval time. The DT then offers an application programming interface (API) for application development that is based on product ids instead of time as the main reference. Furthermore, the twin comes with a graphical user interface that allows for fine-grained insights into the state of the production items (products and machines), a snapshot is shown in Figure 3.

The sensor data attached to the twins is enhanced by the results of so-called "soft sensors" i.e., data obtained from software simulation models. In CAPRI, the solidification sensor (CSS2), temperature sensor (CSS3) and scale sensor (CSS4) are examples of soft sensors. A special role is played by the product tracking system (CSS1); it is the basis for converting time-related data to product-related data, so is essential for the creation of the product digital twins.

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

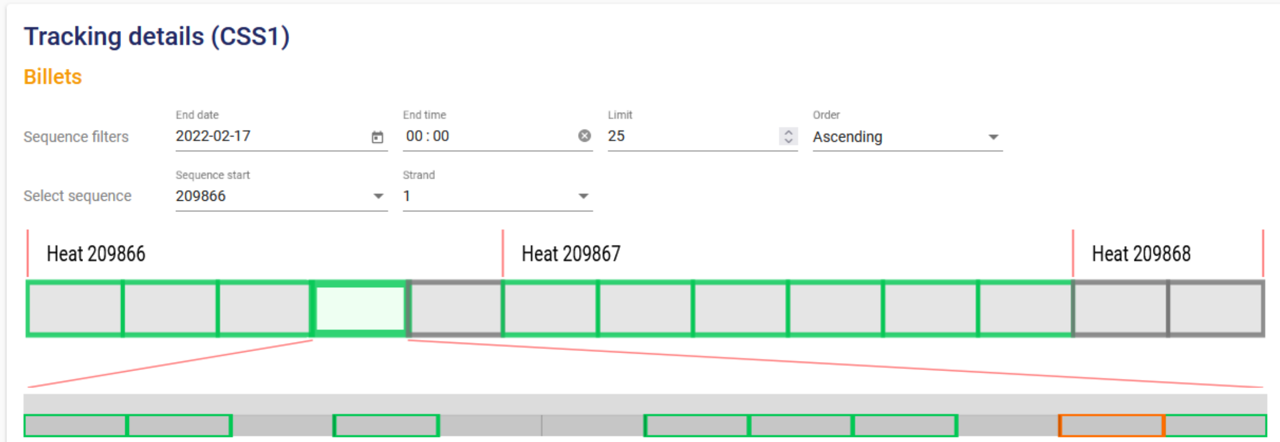

Figure 3: The twin user interface, showing steel billets and bars with a color indicator for their respective tracking status.

The digital twins are enablers for new applications which can access the twin data via the API. One example is the CAPRI risk and anomaly sensor (CSS5), which is a data-driven app that aims to provide a risk estimate for surface quality defects to form, for both steel billets after casting and steel bars after rolling.

The underlying data model of the digital twins is still under development, the finalized laboratory prototypes are based on an ad-hoc model. One special aspect that the model needs to reflect is the tracking uncertainty. Tracking of steel long products during the production is technically quite challenging, and we cannot expect 100% accuracy. For instance, it may occur that a single billet from the casting process is erroneously never identified entering the hot rolling mill or that it is identified more than once, as illustrated in Figure 4. The model must be able to handle these errors and the API must allow the applications to filter for successfully tracked products.

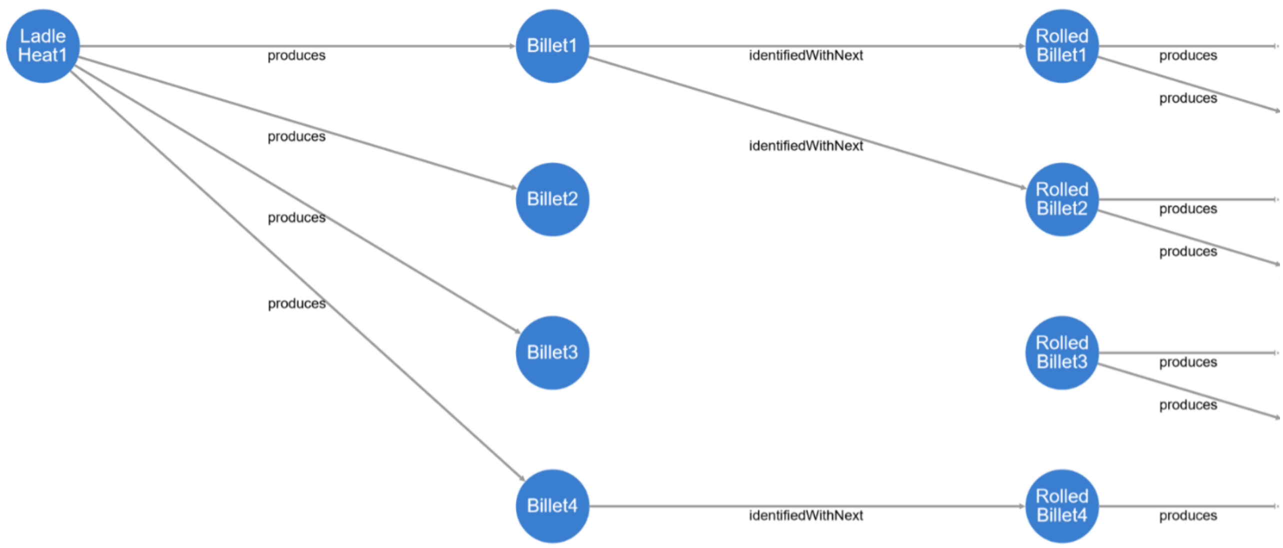

Figure 4: Excerpt of a product graph from the steel production.

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

Multiple billets are produced in the continuous caster from a heat i.e., a batch of liquid steel. Sometime later these billets are processed in the hot rolling mill, but the identity of the billets entering the mill (marked as “RolledBilletX”) comes with an uncertainty. Sometimes billets are uniquely identified, others appear multiple times or not at all.

3.3 Pharma smart knowledge and semantic data model solutions

This section provides a high-level overview of the cognitive sensor solutions CPS3, CPS4, and CPS5 of the pharma use case.

<u>CPS3 – granule moisture estimation</u>: The ConsiGma CTL 25 line, which is part of the pharma use case, contains a fluid bed dryer. This dryer consists of six chambers. They are sequentially filled; the material is dried and then they are sequentially emptied. Only one of the six chambers can be equipped with a process analytical technology (PAT) tool, to measure the granule moisture in that chamber. The granule moisture in the remaining five chambers is not known. The idea of CPS3 is to use a mathematical model of the drying process in order to predict the granule moisture in all of the six chambers. For that purpose, the inlet air temperature and humidity, the outlet air temperature and humidity of the dryer, as well as the measured granule temperature in the individual chambers, is used. Based on the obtained information and the developed mathematical model, the granule moisture in all six chambers is estimated.

The data involved in this solution covers real-time process data that is obtained from the ConsiGma CTL 25 OPC server. The mathematical model is implemented in Matlab/Simulink, running on a dedicated computer, the information about the granule moisture is then used to feed the CPC1 solution (see deliverable D3.3).

<u>CPS4 – dissolution prediction</u>: The dissolution profile of tablets is a critical quality attribute (CQA) and typically determined by offline laboratory analysis. This analysis is destructive, takes a lot of time, and can only be undertaken for a very limited number of tablets. CPS4 should provide a solution that uses real-time process and PAT data, in order to predict the dissolution profile. The data driven model used to predict the profile from the available data has been developed.

<u>CPS5 – fault detection</u>: The aim of CPS5 is the automatic detection of faults and anomalies in a continuous pharmaceutical manufacturing line. Especially in continuous manufacturing lines, all unit operations need to be reliably available and running at the same time. Faults in single pieces of equipment can cause downtime for the entire production line. Therefore, the early detection of anomalies can help in preventing faults and therefore reduce the downtime of the manufacturing line. Data used for CPS5 covers process data that is available from the ConsiGma CTL 25 line, as well as data captured by the other cognitive solutions.

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

4 Smart knowledge and semantic data models cognitive solutions results

At operational level, changes in the production process of cognitive plants can be detected and the plant is able to respond to these dynamic fluctuations by adapting the production to stay within the targets of costs and rate, as well as quality and sustainability. Moreover, high-level business intelligence for the whole plant, focusing on the economic particularities and well-founded decision support systems to abort processing of unviable products or maintenance prediction systems to avoid risky situations can be obtained through smart-knowledge base and semantic models solutions. Different Cognitive Solutions in the three use cases of the CAPRI project have been developed, where models by static but flexible data structures as well as simulated by dynamic digital twins where appropriate have been modelled semantically in an ontology if necessary.

4.1 Asphalt domain

4.1.1 Predictive maintenance of baghouse based on cognitive sensors and expert knowledge (CAO1)

Based on the exploratory data analysis presented in D3.6: Reference Implementation of Cognitive Process Plants and Alignment with other cognitive initiatives, we intend to proceed with:

- The implementation of advanced anomaly detection techniques for identifying any abnormal behaviours of the baghouse.
- Novel machine learning techniques will be performed for predicting the health index of a component of baghouse.

Next, we present a high-level description of these models in real data of ***predictive maintenance of baghouse (CAO1),*** the results of which will be presented in detail in WP4.

The sensor data were collected from April 2021 to November 2021 and concerned information about

- Baghouse filters drop pressure
- Drying drum drop pressure
- Baghouse temperature
- Electric Power blower
  - AMPERAGE Exhauster
  - ELECTRICITE Exhauster
- Information about production orders like: Formula (Recipe) Code (SPA0200), production flow (SPA0400), Formula (Recipe) Name (SPA0300), Final Product Temperature (ºC) (SPA0600), dosification setpoint (SPB0105), bitumen percentage (SPB0110) etc.)
- Maintenance history data

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

4.1.1.1 Anomaly detection model

Generally, *anomaly detection*[4,5] is a significant area in data mining, which identifies events, data points, and/or observations that deviate from normal behaviour. Anomalous data can indicate critical incidents, such as a technical glitch or potential opportunities, for instance a breakdown in a machine.

For the development of an accurate and efficient anomaly detection framework there are four major challenges, which should be taken into consideration:

1. The definition of the normal region, which encompasses every possible normal behaviour, is extremely complex and challenging.
2. Noisy data considerably increases the complexity of problem and the identification of anomaly behaviours.
3. The definition of anomaly is not only domain-specific but can also evolve over time.
4. Getting access to an adequate number of labelled anomalous instances is difficult. Since anomalies are rare occurrences, there are fewer instances available.

CAO1_Image_1.png: Anomaly detection framework for the predictive maintenance of the baghouse

A high-level description of the proposed framework for the predictive maintenance of the baghouse is presented in Figure 5 (CAO1_Image_1.png). In more detail, the proposed AI-based framework takes as input the sensor data from the baghouse as well as the machine specifications (sampling rates, known abnormal behaviours/faults), which are proceed by an intelligent Deep Learning (DL) model. The main idea is the development of a cost function for examining the hypothesis: “if the imported data advocate an abnormal behaviour”. This cost function is usually non-linear and complex, and it is determined by the DL model and its value is utilized for predicting if the machine exhibit an abnormal behaviour. Along this line, the output of the proposed framework are warnings, which indicates that the machine is likely to exhibit a possible abnormal behaviour and alerts, which indicates that the machine may not be working properly.

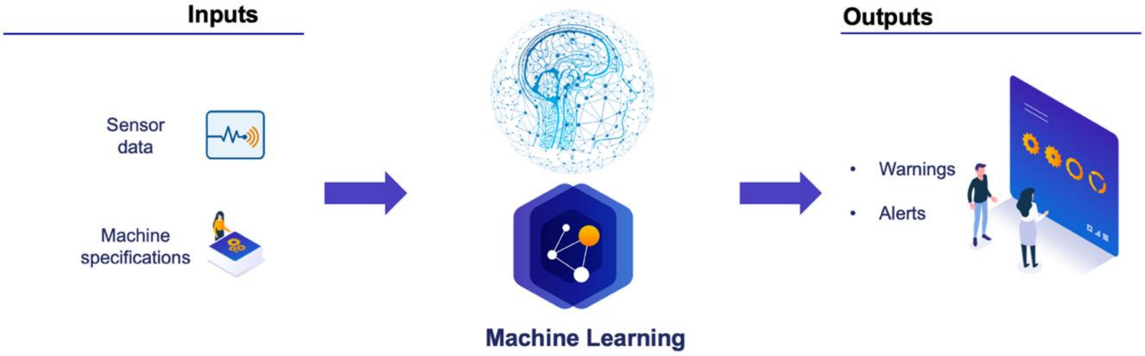

Figure 5: Anomaly detection framework for the predictive maintenance of the baghouse

[4] Pang, G., Cao, L., & Aggarwal, C. (2021, March). Deep learning for anomaly detection: Challenges, methods, and opportunities. In Proceedings of the 14th ACM International

[5] Yao, D., Shu, X., Cheng, L., & Stolfo, S. J. (2017). Anomaly detection as a service: challenges, advances, and opportunities. Synthesis Lectures on Information Security, Privacy, and Trust, 9(3), 1-173.

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

In the context of CAPRI, the developed DL model is a Convolutional-based Auto-Encoder (CAE) model, which is presented in Figure 6. It is worth mentioning that CAE consists of state-of-the-art approaches for anomaly detection tasks, which have been successfully applied in a variety of real-world applications[4,5] .

The CAE model takes as input the sensor and production data and attempts to reconstruct the inputs. The cost function (loss) is defined as the mean square error of the input and the outputs, namely

$$\|\tilde{X} - \hat{X}\|_2$$

where $\tilde{X}$ is the input matrix while $\hat{X}$ is the reconstructed matrix.

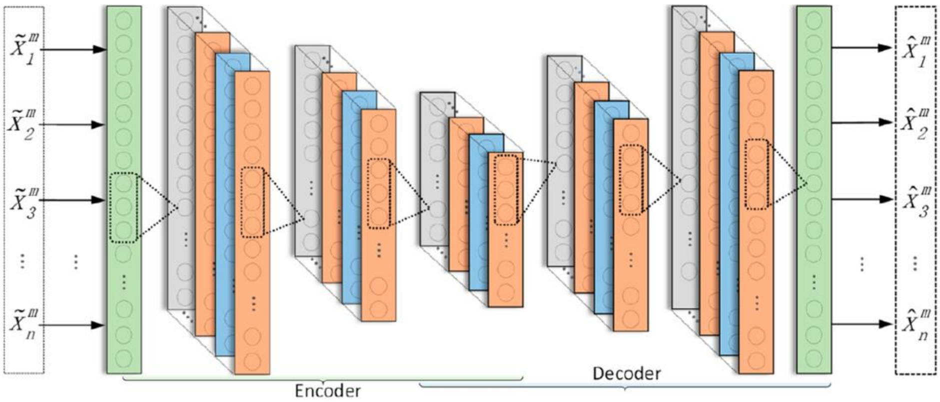

Figure 6. Convolutional Auto-Encoder model

Based on the value of the loss, the input data can be characterized as “Normal”, “Warning”, “Critical” based on two pre-defined thresholds T_{WARN} and T_{CRIT} (also called as *decision thresholds*) with $T_{WARN} < T_{CRIT}$. More specifically,

- If the value of the cost function is less T_{WARN}, then the input data are characterized as “Normal”, which indicates that the machine is predicted to be working well.
- If the value of the cost function is between the values T_{WARN} and T_{CRIT} then the input data are characterized as “Warning”, which indicates that the machine is likely to exhibit a possible abnormal behaviour.
- If the value of the cost function is greater then T_{CRIT}, then the input data are characterized as “Critical”, which indicates that the machine may not be working well.

Notice that these pre-defined thresholds are defined during the training process of the ML model based on statistical analysis.

CAO1_Image_2.png: Example from the application of the anomaly detection framework.

Figure 6 (CAO1_Image_2.png) reports an example from the application of the proposed framework for the identification of possible anomalies. The horizontal axis represents the time while the vertical axis represents the value of the cost function. With the yellow and red color are denoted the threshold T_{WARN} and T_{CRIT}, respectively. The interpretation of the Figure 7 illustrates that between 11:20am and 11:35am, the model’s prediction is “Warning”, which suggests that the machine is likely not to be working well.

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

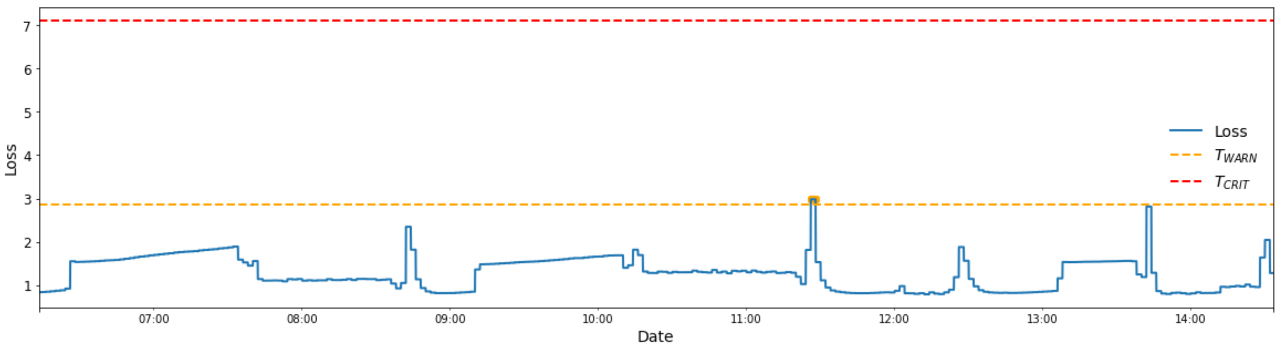

Figure 7: Example from the application of the anomaly detection framework

Regarding the following steps, we intend to proceed with

- The application of unsupervised feature selection techniques for improving and possible optimize the performance of the developed CAE model
- Exploit maintenance information, regarding known faults and/or abnormal behaviors of the baghouse in order to evaluate the performance of the developed model.
- Conduct a detailed statistical analysis of the predicted anomalies in order to provide a deep insight to the reliability of the model.

4.1.1.2 Remaining useful life

The *remaining useful life* (RUL)[6] of a machine consists of the expected life or usage time remaining before the machine requires replacement or repair. The accurate prediction of RUL from system data generally considered one the main goals of predictive maintenance.

The term *lifetime* or *usage time* refers to the life of the machine defined in terms of whatever utilized quantity for measuring the machine's life. Notice that the units of lifetime can be quantities such as the distance travelled (miles), fuel consumed (gallons), repetition cycles performed, or time since the start of operation (days). Similarly, *time evolution* may mean the evolution of a value with any such quantity.

Typically, the RUL of a machine is estimated by developing a prediction model, which can perform the estimation based on the time evolution or statistical properties of condition indicator values, such as:

- A model, which compares the time evolution of a condition indicator to measured or simulated time series from systems that ran to failure. Such a model can compute the most likely time-to-failure of the current machine.
- A model, which fits the time evolution of a condition indicator and predicts how long it will be before the condition indicator crosses some threshold value indicative of a fault condition.

[6] Hu, Y., Liu, S., Lu, H., & Zhang, H. (2019). Remaining useful life model and assessment of mechanical products: a brief review and a note on the state space model method. Chinese Journal of Mechanical Engineering, 32(1), 1-20.

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

Predictions from such models are statistical estimates with associated uncertainty. They provide a probability distribution of the RUL of the test machine.

CAO1_Image_3.png: Remaining of Useful Life framework
A high-level description of the RUL framework, which will be implemented in the context of CAPRI is presented in Figure 8 (CAO1_Image_3.png). More analytically, the framework that as input the sensor data as well as maintenance info (breakdown durations of machine, date, and time of breakdown etc.), which be processed and transformed in order to be utilized for fitting and training an efficient ML model. The output of the model will be the estimated time until next failure of a component of baghouse.

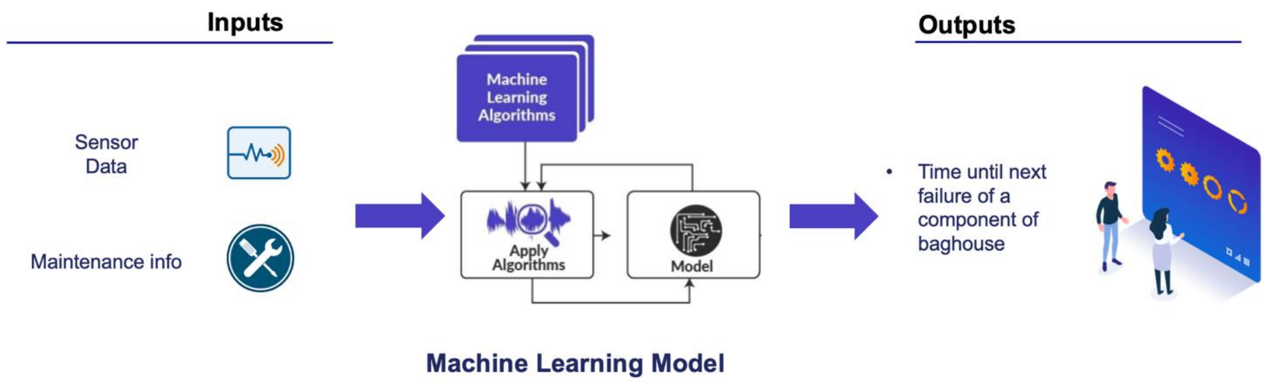

Figure 8: Remaining of Useful Life framework

Figure 9 reports an example from the application of the proposed framework for the estimation of remaining useful life. The horizontal axis represents the date while the vertical axis represents the remaining useful life (in days). With the green and yellow line are denoted the actual and the predicted by the model remaining useful life.

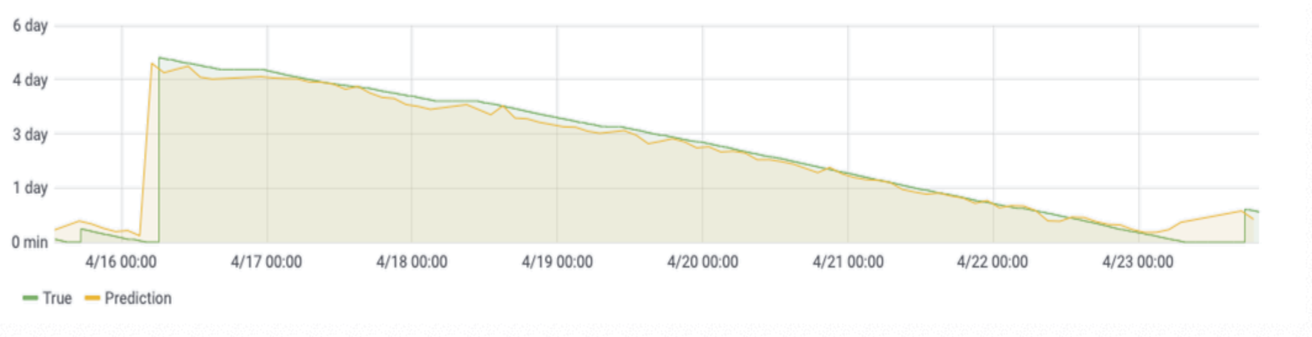

Figure 9: Example from the application of the remaining useful life framework

In the future, we will proceed with the implementation of proposed framework exploiting any collected/retrieved maintenance information and evaluate its performance.

4.2 Steel domain

4.2.1 Digital Twins for steel long products (CSO1)

In the CAPRI project, a significant amount of production data has been collected and fed into the digital twin. This includes the steel composition, casting timeseries data, such as the casting speed and tundish temperature (CSS2 input), billet tracking data (CSS1), various datapoints from the hot rolling mill, such as temperature measurements (CSS3 input), rolling forces, etc., and steel bar surface quality information (CSS5 labels). In addition, results of the sensors CSS2-CSS5 enter the twins.

Since the bar tracking was not yet operational at the time of writing, the focus was on so-called one-billet orders, special customer orders consisting of a single billet, because certain aggregated tracking information could be provided for these billets and their resulting bars even without the new bar tracking system. The data collection period covered the greater part of the year 2021 and the first month of 2022. Hot rolling data is only available for two months in 2021 plus the 2022 data. After removal of billets with incomplete data our initial dataset contains some 400 billets with process data and labels.

The data model under development is unpublished as yet and based on the SAREF4INMA[7] (SAREF for industry and manufacturing) ontology. Furthermore, it follows the FIWARE modelling approach[8] of representing the physical objects in terms of a property graph, where nodes of the graph represent entities (e.g. products and machines), and edges represent relationships. Both nodes and edges can have properties and they are typed, which means that the property names and value types they carry are defined by some class structure. An example is shown in Figure 10. As explained in the introduction, the tracking uncertainty implies that we cannot always track a given item unambiguously through the processing chain. For this reason, the billet is duplicated, as a cast billet and a rolled billet, and a relationship “identifiedWith” exists between the two, which is 1-1 in the case of successful and unique tracking, but may be n-m in general.

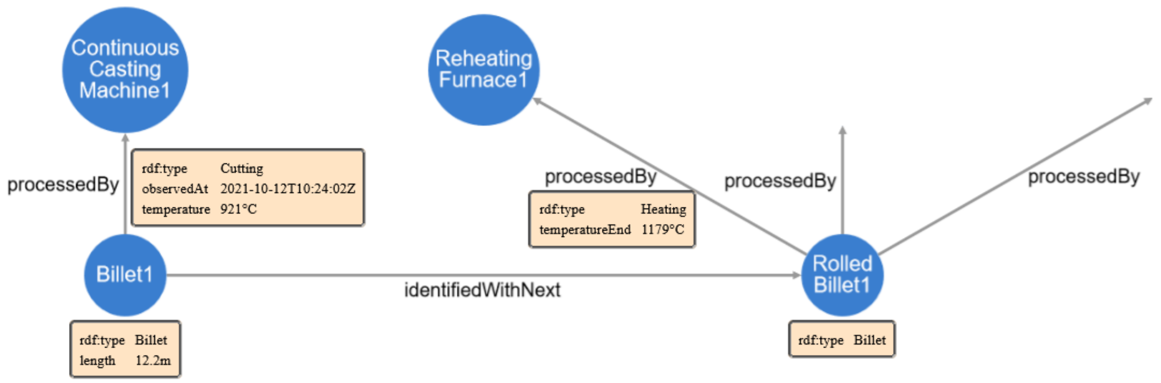

Figure 10: Sample property graph showing two products (billets) and two machines (caster and reheating furnace), along with some relationships and properties associated to both entities and relationships

[7] https://saref.etsi.org/saref4inma
[8] https://www.etsi.org/committee/cim

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

The preliminary high-level class structure is shown in Figure 11. It is based on the SAREF ontology with the SAREF4INMA extension, plus some additions and small adjustments. Furthermore, we aim to adhere to the NGSI-LD cross-domain ontology rules, which also leads to some deviations from the SAREF standard. The most important classes are

- **ProductionEquipment**: machinery used in the production process.
- **ProductionResource**: everything from raw material through intermediate products to the end products, including waste. This is further subdivided into individual items, batches of items and batches of material (classes **Item**, **ItemBatch** and **MaterialBatch**).
- **Factory** and **BuildingSpace**: spatial categories. BuildingSpace is further subdivided into **Site** and **Area**.
- **FeatureOfInterest**: everything that has measurable quantities. These measurable quantities are themselves represented by the **Property** class. In the original SAREF4INMA specification, only *ProductionResources* are *FeaturesOfInterest* (strictly speaking, the subclasses of ProductionResource, which itself does not exist in SAREF4INMA), whereas *ProductionEquipment* is not, implying that all measurements have to be assigned to (intermediate) products or material. In our case, measurements are often assigned primarily to the production equipment, even if we would ultimately like to link them to the products as well.

ProductionEquipment can be part of a larger unit, indicated by the **isPartOf** relationship and its inverse, the **consistsOf** relation.

The connection between *ProductionResources* and the *ProductionEquipment* used in their making is given by the **processedBy** relationship (replaces SAREF4INMA *needsEquipment*). Thus every product carries a relation to the machines that have been used in its processing, but not viceversa. We assume the *processedBy* relationship to be ordered. Every *processedBy* relation is assigned a type, namely a subtype of the *Transformation* class, such as *Heating*, *Cooling*, *Cutting* or *ShapeTransformation*. These will be listed below.

Connections between products are modeled by the **madeFrom** relationship and its inverse, the **produces** relationship. For instance, these may connect a billet and the bars cut from it after rolling.

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

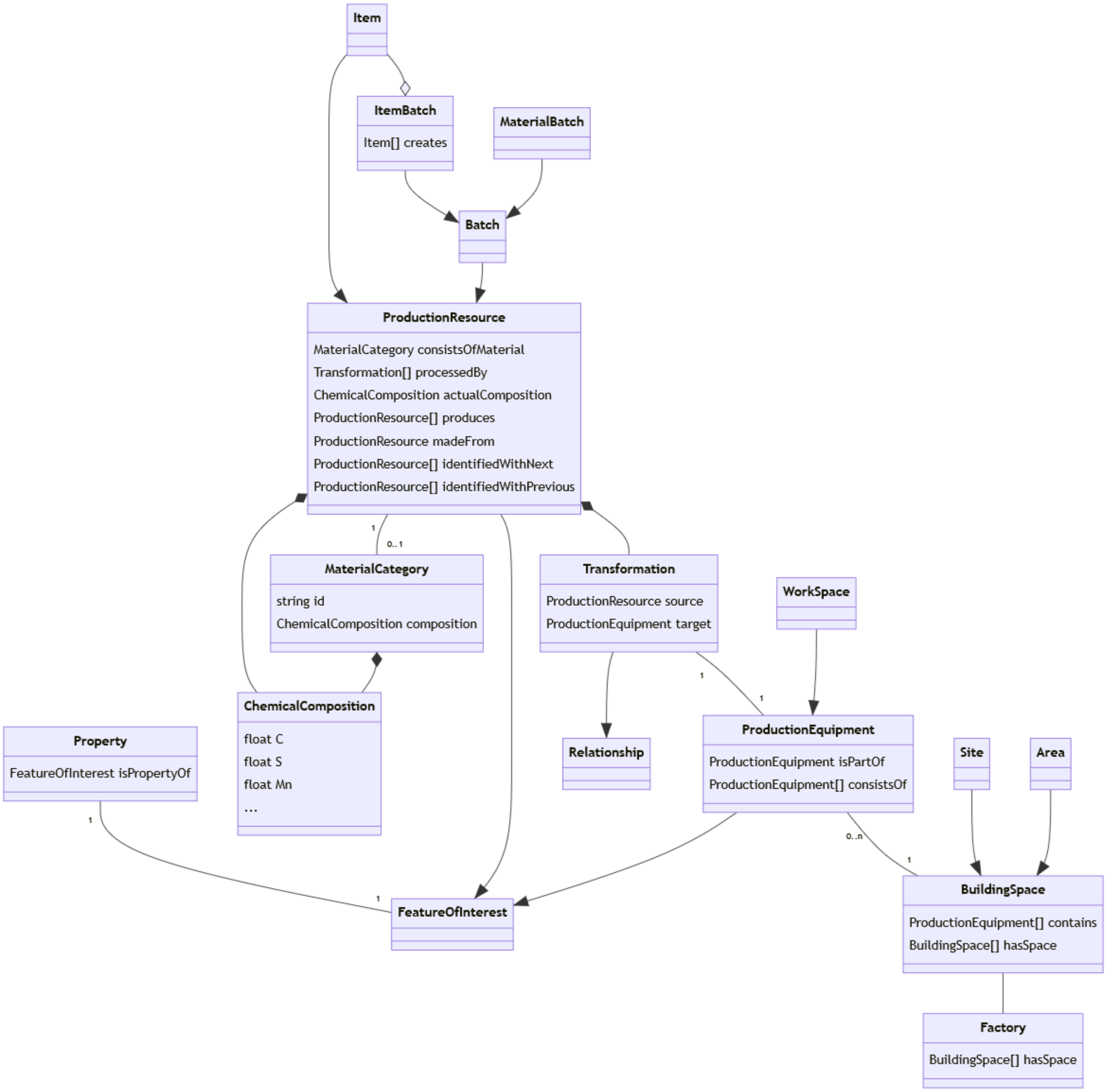

Figure 11: High-level class structure for the steel domain, based mainly on the SAREF4INMA ontology (https://saref.etsi.org/saref4inma/v1.1.2/).

CSO1_video_1.mp4: Digital twin concepts and user interface

The video explains the concept of the digital twins and provides an overview of the graphical user interface. The latter visualizes the data from the different data sources and in particular the billets and bars with their respective tracking status.

Since the sensor data from the automation system is usually time-related, a conversion or an interpretation of the data needs to be done by the twin, either upon ingestion or on the fly at retrieval time. The DT then offers an API for application development that is based on product ids instead of time as the main reference. The conversion of the data is based on the tracking data provided by CSS1.

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

CSO1_API_1.pdf: Data access application programming interface

The API of the product twins needs to provide access to three different types of data:

1. Instantaneous data
2. Structured data
3. Timeseries data

Instantaneous data reflects a currently valid sensor value, for instance. Structured data contains information that does not normally change with time, for instance the chemical composition of a heat i.e., a batch of liquid steel. Examples of timeseries data are historical recordings of sensor values. A generic specification has been developed that covers these three cases. It is explained in the referenced document.

4.3 Pharma domain

4.3.1 Cognitive Operation Concept (CPO1)

The goal of CPO1 is to develop an algorithm that provides suggestions for the operator in order to avoid faults/downtimes. To cope with diverse realistic scenarios observed in the ConsiGma 25™ (GEA) operation, this cognitive solution CPO1 has been divided into three algorithms.

Algorithm 1 – Dissolution

This algorithm provides the main compaction force (MCF) suggestion for a rotary tablet press to the operator. The dissolution model developed in CPS4 is utilized to predict the tablet dissolution properties of a product key (PK) entering the tablet press. The material stream is split into PKs in the fluidized bed dryer, which consists of six chambers. The material dried in one specific chamber is defined as a PK. The main factors influencing the dissolution are M1 of the granule size distribution and the granule moisture, expressed in terms of loss on drying (LOD), which are acquired via CPS2 and CPS3, respectively. They are correlated to the corresponding PK by means of the residence time distribution (RTD) model developed in CPC1. For the prediction purpose, the currently set MCF is used, and a deviation between predicted and reference dissolution curve is calculated. In case this deviation does not exceed a certain threshold, the following operator message appears: ***Predicted dissolution is OK. Please continue with the main compaction force of X kN.*** Otherwise, an optimization problem is solved to calculate the MCF such that this deviation is compensated. Finally, the optimal MCF is utilized to predict dissolution and to re-calculate the deviation. If the algorithm manages to reduce the deviation below the specified threshold, the following operator message appears: ***Predicted dissolution deviation. Please adjust the main compaction force to X kN.*** Otherwise, the algorithm provides discharge suggestions to the operator and the following message appears: ***Predicted dissolution deviation. Please discharge this PK.***

Algorithm inputs: M1 (CPS2) and LOD (CPS3)

Algorithm parameters: dissolution model (CPS4), RTD model (CPC1)

Algorithm outputs: MCF and operator message

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

The zip archive **CPO1_App_and_data1.zip** contains the implementation of the algorithm in Python and a sample dataset (subfolder /excel) to be used for test purposes.

Besides this algorithm, two other algorithms are currently being developed, too:

Algorithm 2- Tablet press hopper
This algorithm provides tableting speed suggestions to the operator. Initially, the weight of the product key (PK) entering the tablet press is measured via a scale and compared to the ideal PK weight. In case the difference exceeds a certain threshold, the tableting speed is recalculated, and the following operator message appears: ***PK weight exceeds the valid range, please adjust tableting speed to X rpm***. Additionally, the hopper fill level in the tablet press is measured, and utilized to generate operator suggestions. In case the deviation of the hopper level from the reference value exceeds a certain threshold, the following operator message appears: ***Hopper level exceeds the valid range, please adjust tableting speed to X rpm***.

Algorithm inputs: scale signal and hopper level signal

Algorithm outputs: tableting speed and operator message

Algorithm 3
This algorithm should provide operator suggestions to check for potential equipment faults in the granulation line, based on the process models developed in CPS1 and CPS2. For this algorithm, it is distinguished between two operation scenarios (CPS1 Raman OR CPS2 Parsum probe installed at the granulator outlet). In the first operation scenario, via CPS1 the predicted liquid-to-solid (LS) ratio is compared to the LS set point. In case the difference exceeds a certain threshold, an operator message appears: ***Please check for potential equipment faults (solid feeders and liquid pump) in the granulation line!***

Algorithm input: liquid to solids ratio obtained via CPS1 (LS_{CPS1}), setpoint of liquid to solids ratio (LS_{SP})

Algorithm output: operator message

In the second operation scenario, the particle size distribution (PSD) soft-sensor outputs (the soft-sensor is utilizing LS_{SP} as model input to predict $\widehat{M}_1 \dots \widehat{M}_4$) are compared to the CPS2 outputs (Parsum measurement $M_1 \dots M_4$). If the difference exceeds a threshold, the same operator message appears: ***Please check for potential equipment faults (solid feeders and liquid pump) in the granulation line!***

Algorithm inputs: $\widehat{M}_1 \dots \widehat{M}_4$ and $M_1 \dots M_4$ (CPS2)

Algorithm parameters: operator message

Table 3: Contents of CPO1 zip archive, CPO1_App_and_data1.zip

Filename	Contents
CPO1_App1.py	Pyhton script that implements algorithm 1
description.txt	Explanation of the python script CPO1_App1.py
Excel/A3_DOK_disso_operating_range.xlsx	operating ranges required for normalisation of individual model inputs
excel/A3_DOK_disso_training_data.xlsx	columns correspond to captured model inputs:

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

	main compaction force (MCF), PSD moment1 (M1) and loss-on-drying (LOD)

The DMP (data management plan) associated to this CS data is included in the file CPO1_DMP_CPO1_App1_and_data1_V1.xlsx

4.3.2 Cognitive sensor for product moisture (CPS3)

The proposed solution is a soft sensor used to predict the granule moisture during the drying process in the ConsiGma CTL 25 segmented fluid bed dryer. The dryer consists of six chambers, each of them equipped with a temperature sensor for measuring the granules' temperature. One chamber supports the *measurement* of granule moisture by means of near infrared (NIR) spectroscopy. The remaining five chambers do not have that option. Therefore, the process data available via the ConsiGma CTL 25 OPC server is used to feed a mathematical model of the dryer, which in turn estimates the granule moisture in all the six chambers. The model is based on energy and mass balance equations.

Figure 12, presents the process data generated from the ConsiGma CTL 25 line and fed into the soft sensor. Data is obtained via OPC at a sampling rate of 1 second. The quantities on the left show values that are related to the entire dryer (i.e., they are not specific to the individual chamber: e.g., there is only one common inlet air stream with temperature T_a^i and water content x_a^i), whereas the value on the right (granule temperature T_g) is available for each chamber.

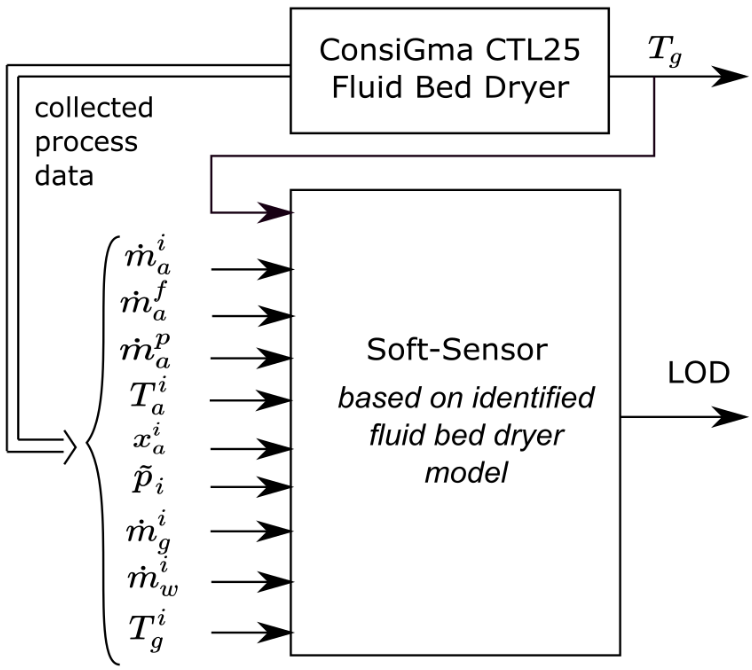

Figure 12: Input and output data of the CPS3 solution

The soft sensor has been implemented in Matlab/Simulink. The block diagram is depicted in Figure 13 and Figure 14. Besides the necessity of blocks based on algebraic equations required for

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

modelling the air mixing, the material and energy transfer between granules, air inside the dryer and ambient air is modelled by means of ordinary differential equations (ODEs). For each chamber, a set of ODEs has been implemented. A variable step solver is used for numerically solving the model equations.

An example data set that has been created using this simulation model is provided in **CPS3_data1.zip**. It contains the files noted in Table 4.

The DMP associated to this CS data is included in the file CPS3_DMP_CPS3_data1_V1.xlsx.

Table 4: Contents of CPS3 zip archive, CPS3_data1.zip

Filename	Contents
A3_DOK_CPS3_process_data.xlsx	Sample dataset
Explanation.txt	Explanation of the variables provided in A3_DOK_CPS3_process_data.xlsx

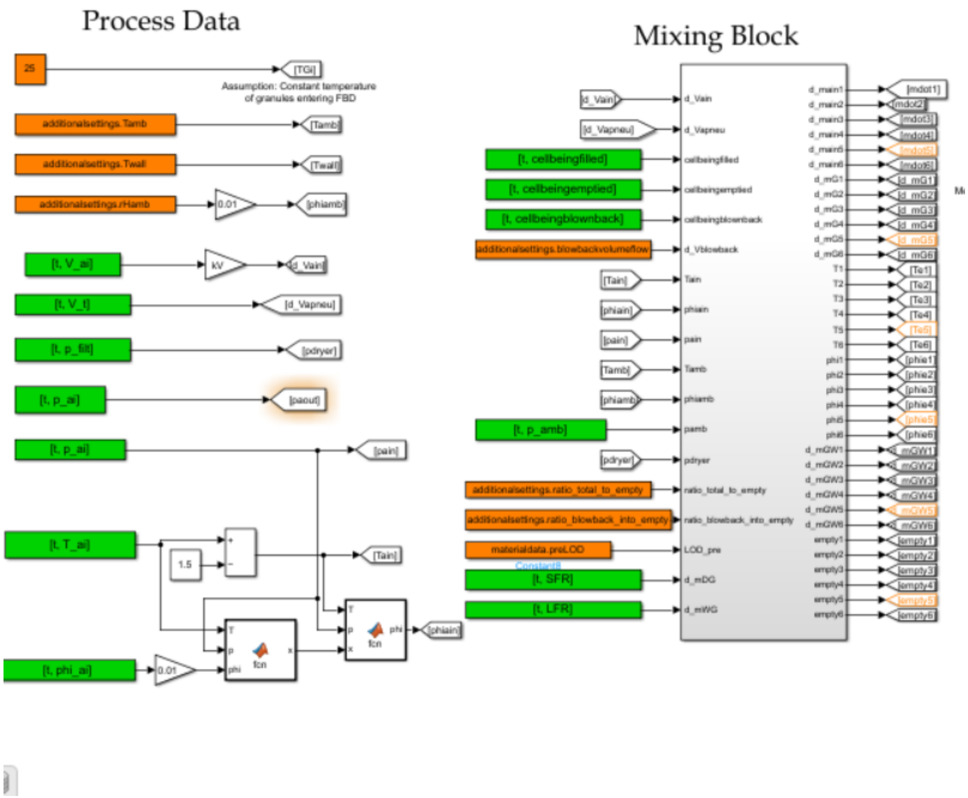

Figure 13: Simulink block for air mixing

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

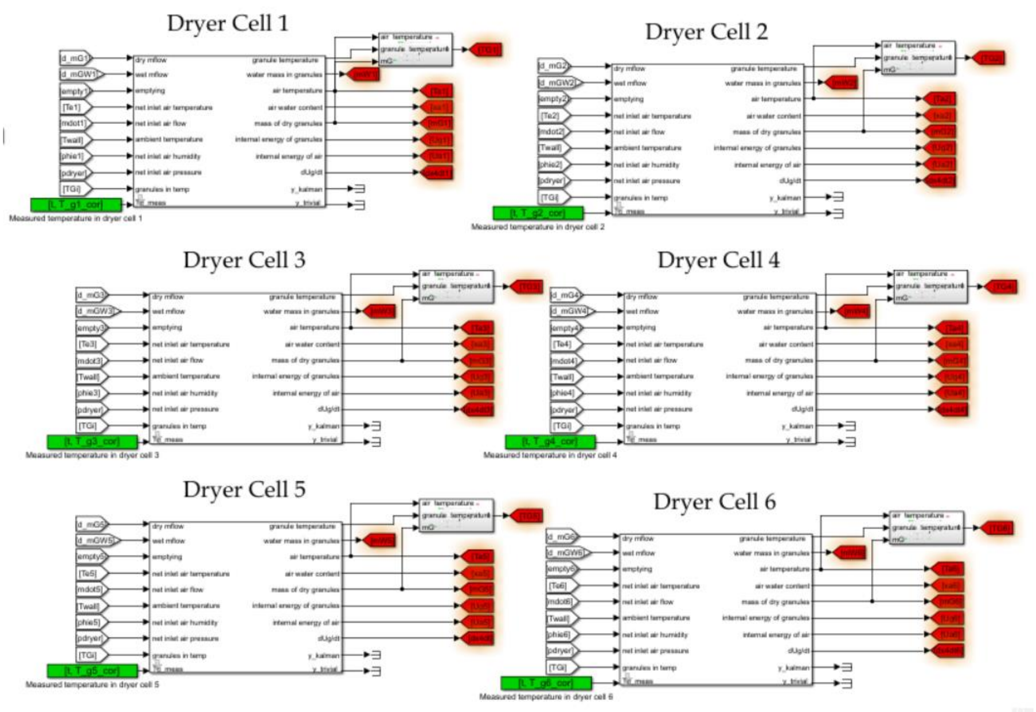

Figure 14: Simulink blocks for solving energy and mass balances in the respective dryer cells

4.3.3 Cognitive prediction of dissolution (CPS4)

The dissolution profile of a tablet is influenced by the process settings during production. A data driven model has been developed in order to predict the dissolution profile of tablets from the process settings, PAT and cognitive sensor data. Granulator and dryer process settings were varied in order to capture the effect of particle size distribution at the granulator, and loss-on-drying (LOD) after the dryer, on the dissolution profile. Additionally, the effect of the main compaction force (MCF) at the tablet press was investigated. The final inputs to the model are the first moment of the particle size distribution M1, LOD and MCF. They are used to predict parameters of a dissolution curve model.

Figure 15 shows the blocks required for predicting the dissolution profile. The LOD data, as well as the output of CPS2, need to be aligned in a timely manner to the compaction force of the individual tablets. For that purpose, a material tracking model was developed (see CPC1 in deliverable D3.3). The test data provided within the attached zip-file A3_EXE_CPS4_20220222.zip contains steady state operation data, i.e., the process settings remained constant throughout the trial. Due to this, the RTD model used for timely alignment is not needed.

The zip-archive **CPS4_App_and_Data_1.zip** contains the files described in Table 5.

Table 5: Contents of CPS4 zip-archive, CPS4_App_and_Data_1.zip

excel/A3_DOK_CPS4_model_data.xlsx	*Example data to be used for test purposes with respect to the data-driven correlation model (see* Figure 15*)*
excel/A3_DOK_CPS4_model_parameters.xlsx	*PLS model coefficients for model 1 (BetaPLS 1 for T_1) and model 2 (BetaPLS 2 for T_2)*

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

excel/A3_DOK_CPS4_operating_range.xlsx	*Operating range of the model inputs*
description.txt	*Textual description of the files contained in the zip-archive*
A3_exe_CPS4_test_V01.py	*Python script to perform model prediction (data-driven correlation model in* Figure 15*)*
A3_exe_CPS4_test_V02.py	*Python script to perform model prediction (data-driven correlation model in* Figure 15*)*

The algorithm contains two parts: in the first part, the dissolution curve itself is described by two parameters T_1 and T_2 (see right block in Figure 15); the second part translates the mentioned process settings into the parameters T_1 and T_2 (see left block in Figure 15).

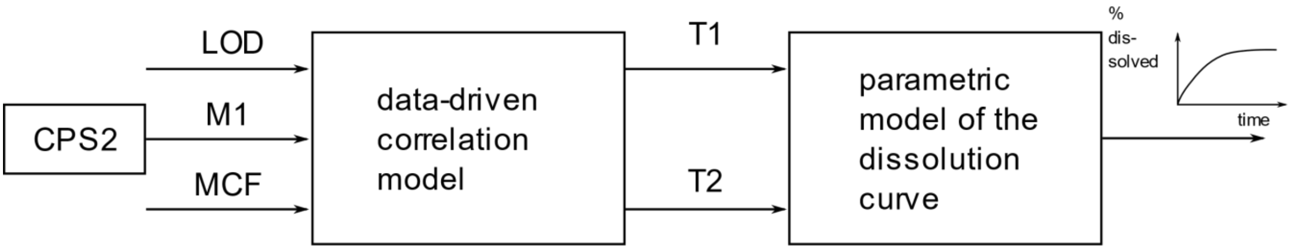

Figure 15: Modelling approach of CPS4

The parametric model used for describing the dissolution profile (concentration c as a function of time t) is given by

$$c(t)[\%] = 100\ \% \left(1 - \frac{T_1}{T_1 - T_2}\, e^{-\frac{t}{T_1}} + \frac{T_2}{T_1 - T_2} e^{-\frac{t}{T_2}}\right).$$

The parameters T_1 and T_2 were identified for a set of measured dissolution profiles.

As a next step, a correlation model was trained to predict the parameters T_1 and T_2 from the process data, i.e., M1, LOD and MCF. For that purpose, many dissolution profiles were modelled and the corresponding process data were recorded. The models (one for predicting T_1, one for predicting T_2) were trained by means of a partial least squares (PLS) regression method. Figure 16 shows one example comparing the measured dissolution curve to the parametric and PLS modelled dissolution curve.

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

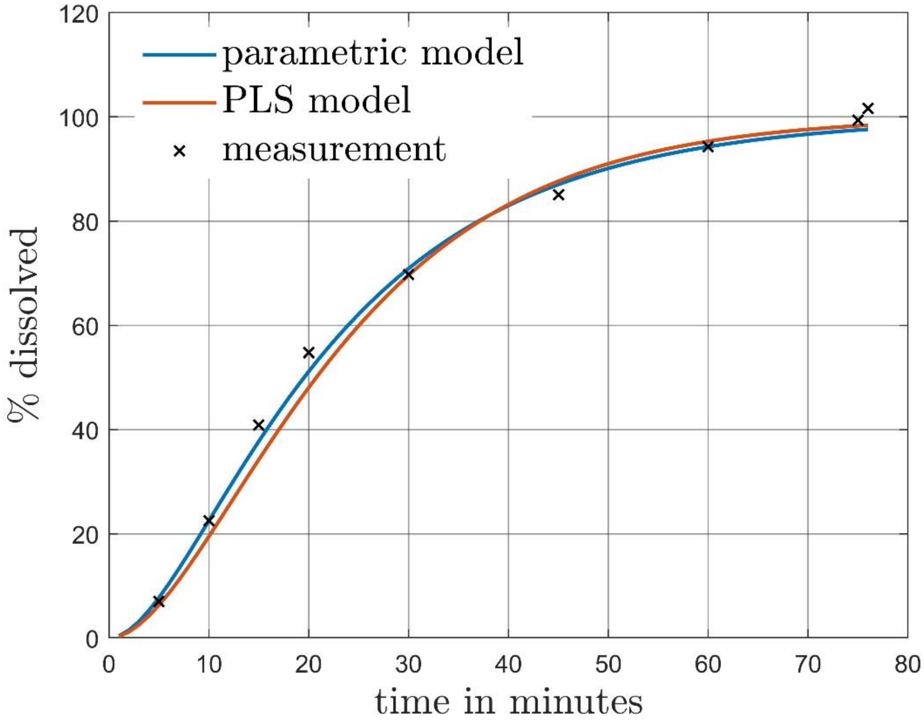

Figure 16: Measured dissolution profile (black crosses), parametric model output (blue), and PLS model output (red)

In order to further enhance the correlation model described above, additional experimental data was obtained offline. This data was used to develop preliminary models and to evaluate the capability of a novel sensor, as an additional in-line cognitive component. A novel image based soft-sensor concept emerged as highly promising. Soft-sensor measurements were taken offline for 42 tablets and compared to parameterised dissolution profiles. Based on this, a machine learning based correlation model was developed by using the graph-based method k-nearest-neighbours. In three validation examples, the predicted dissolution profile was then compared to the fitted curve in Figure 17.

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

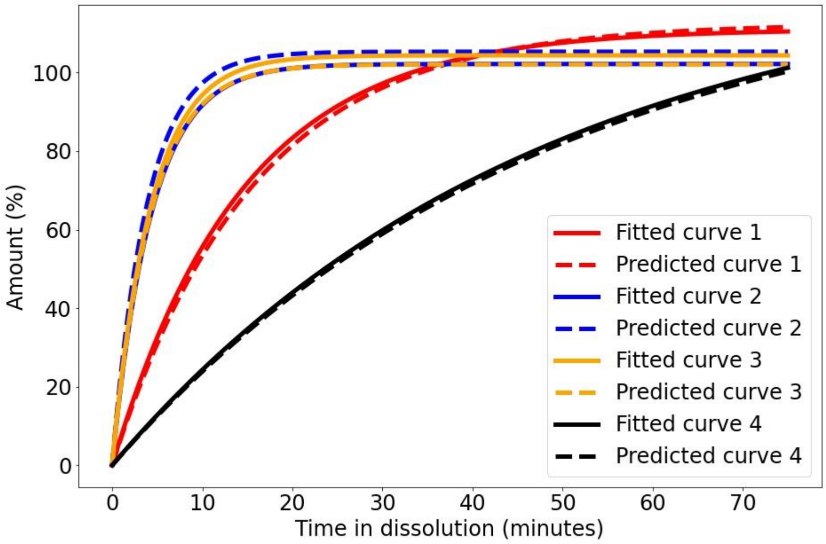

Figure 17: Predicted vs. fitted dissolution profiles.

In the next step, the developed soft-sensor will be transferred for in-line operation and connected to the cognitive solution.

The DMP (data management plan) associated to this CS data is included in the file CPS4_DMP_CPS4_App_and_data1_V1.xlsx.

4.3.4 Cognitive sensor for fault detection (CPS5)

The purpose of CPS5 is to detect abnormalities in a continuous manufacturing line. The proposed solution takes real-time information from the ConsiGma CTL 25 manufacturing line. It is not a single sensor value that is used, but rather multiple sources of information are taken into account by the CPS5 algorithm.

The real time data processed by CPS5 includes temperature, pressure, volume flow and weight data. It is provided by the OPC server of the ConsiGma CTL 25 line. In addition to the data provided by the processing line, information about faults needs to be converted into a standard format needed for training the proposed algorithms. The JSON format was chosen to describe these faults. An example of two abnormalities described by the json format is shown below. It includes the timestamp, the type of fault (C_WRN_XFE06021_TOL), the unit operation(s) involved (dryer), the process data suitable for detecting the fault (XFT_volume_flow_air_dryer) and a description of the anomality (Air flow deviation warning).


```
[
	{
		"timestamp":"2021-03-24 09:33:31",
		"tagname":"C_WRN_XFE06021_TOL",
		"stage":[
			"Dryer"
		],
		"parameters":[
			"XFT_volume_flow_air_dryer"
		],
		"cause":"Air flow deviation warning"
```


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

```
        },
        {
                "timestamp":"2021-03-24 10:00:23",
                "tagname":"C_WRN_XFE07011_TOL",
                "stage":[
                        "TSG",
                        "Dryer"
                ],
                "parameters":[
                        "XFT_volume_flow_air_granule_transport"
                ],
                "cause":"Wet transfer flow deviation warning"
        }
]
```

We provided multivariate anomaly detection based on Ensemble learning. The ensemble consists of multivariate control charts MEWMA and Hoteling and PCA-based analysis. The algorithm creates process behaviour map (outlier view), whereas the periods with anomalies are coloured in red. This enables a proper contextualization of individual outliers. Moreover, the method provides root causes analysis, determining which parameter(s) are responsible for the detected anomalies (outliers). In **CPS5_Data_1.zip**, a sample dataset is provided.

Table 6: Contents of CPS5 zip-archive CPS5_Data_1.zip

20210811_CAPRI_089_Day1.csv	*Sample dataset of one trial-day*
20210813_CAPRI_089_Day2.csv	*Sample dataset of one trial-day*
20210816_CAPRI_089_Day3.csv	*Sample dataset of one trial-day*
20210817_CAPRI_089_Day4.csv	*Sample dataset of one trial-day*
20210818_CAPRI_089_Day5.csv	*Sample dataset of one trial-day*
BCM_Alarm-file_20210811.json	*Information on events in JSON format, related to 20210811_CAPRI_089_Day1.csv data*
BCM_Alarm-file_20210813.json	*Information on events in JSON format, related to 20210813_CAPRI_089_Day2.csv data*
BCM_Alarm-file_20210816.json	*Information on events in JSON format, related to 20210816_CAPRI_089_Day3.csv data*
BCM_Alarm-file_20210817.json	*Information on events in JSON format, related to 20210817_CAPRI_089_Day4.csv data*
BCM_Alarm-file_20210818.json	*Information on events in JSON format, related to 20210818_CAPRI_089_Day5.csv data*

The DMP (data management plan) associated to this CS data is included in the file CPS5_DMP_CPS5_Data1_V1.xlsx

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

5 Conclusions and Next Steps

D3.4 – "CAPRI Smart knowledge and semantic data models", is a deliverable of type OTHER that, together with D3.2, D3.3 and D3.5, contributes to achievement of milestone 'MS4' related to technology validation of different cognitive solutions at M24. The document describes the concrete achievements in the development of the standalone smart cognitive solutions at Operational level. Main practical results have been detailed for each use domain (Asphalt, Steel and Pharma) and cognitive solution, providing designs, algorithms, videos, manuals, source codes, data format and data samples showcasing the implementation done in laboratory activities (WP3 activity) before actual industrial implementation.

The output of WP3, will be integrated with the CAP platform to satisfy the needs of the three CAPRI domains (i.e., Asphalt, Steel, and Pharma) supporting all use cases and covering the entire data life 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 in WP4. After the integration of the cognitive modules, the platform will be tested and tuned, thus feeding the validation scenarios in the three plants, to be addressed in WP5 through two iterations. On the other hand, a toolbox of cognitive solutions for sensing, control, operation and planning will be developed to help the adoption of the CAP for batch, continuous and hybrid process industry plants. The final validation will take place in WP5, addressing manufacturing challenges in industrial operational environments of the three chosen process sectors, and providing useful feedbacks and lessons learned.

Next steps will be in WP4 where cognitive solutions developed within WP3 will be integrated with the CAP following a modular and iterative approach, to provide a holistic solution, managing the cognitive functions embedded in each cognitive solution. At the same time, the integration will enable a higher level of intelligence exploiting the vertical integration with other processing modules, at both edge and cloud level, delivering the cognitive applications towards the appropriate user role (e.g. planners, managers, workers), with first deliverables due by M28.

The progression in the maturity of the cognitive solutions will enable an in-depth analysis in WP7, for the shaping of the exploitation opportunities of the different CAP layers and the CAP platform as a whole, as well as opportunities for replicating cognitive solution in other sectors of process industry.