



D3.2 - Data Management and Harmonization Plan

WP3: Real-Time Water Efficiency Monitoring Platform for the Process Industry

June 2023

Authors: Dimitris Apostolou (ICCS), Babis Magoutas (ICCS), Efthimios Bothos (ICCS), Dimitra Pournara (ICCS), Gregoris Mentzas (ICCS), Aitor Corchero Rodriguez (EUT), Robert Sanfeliu Prat (EUT), Janelcy Alferes Castano (VITO)



Document Information

| | | | |
|-------------------------------|---|----------------------------|-----------------------|
| GRANT AGREEMENT NUMBER | 958396 | ACRONYM | AquaSPICE |
| FULL TITLE | Advancing Sustainability of Process Industries through Digital and Circular Water Use Innovations | | |
| START DATE | 1 st December 2020 | DURATION | 48 months |
| PROJECT URL | www.AquaSPICE.eu | | |
| DELIVERABLE | D3.2 – Data Management and Harmonization Plan | | |
| WORK PACKAGE | WP3 – Real-Time Water Efficiency Monitoring Platform for the Process Industry | | |
| DATE OF DELIVERY | CONTRACTUAL | 03/2022 | ACTUAL 06/2023 |
| NATURE | Report | DISSEMINATION LEVEL | Public |
| LEAD BENEFICIARY | ICCS | | |
| RESPONSIBLE AUTHOR | Dimitris Apostolou | | |
| CONTRIBUTIONS FROM | Babis Magoutas (ICCS), Efthimios Bothos (ICCS), Dimitra Pournara (ICCS), Gregoris Mentzas (ICCS), Aitor Corchero Rodriguez (EUT), Robert Sanfeliu Prat (EUT), Janelcy Alferes Castano (VITO) | | |
| ABSTRACT | This deliverable describes a seamless data management & harmonisation process for effective integration of heterogeneous data sources, varying in: i) content (e.g. data from physical sensors and information systems); ii) quality (e.g. from precise well-defined data produced by static, trusted knowledge sources to dynamic data from external sources); and iii) speed of update (e.g. updated every few hours to highly dynamic, real-time, streaming data) will be defined. | | |

Document History

| VERSION | ISSUE DATE | STAGE | DESCRIPTION | CONTRIBUTOR |
|---------|---------------|--------------|--|--|
| V0.5 | 22 March 2022 | Draft | First version of the document, prior to internal AquaSPICE members review. | Dimitris Apostolou, Babis Magoutas, Efthimios Bothos, Aitor Corchero Rodriguez, Robert Sanfeliu Prat |
| V0.5 | 01 April 2022 | Review | Comment and suggestions | Klemen Kenda |
| V1.0 | 12 April 2022 | Final | Final version | Dimitris Apostolou |
| V2.0 | 30 April 2023 | Resubmission | Revised version considering CS#4 as void case | Dimitris Apostolou |

Disclaimer

Any dissemination of results reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains.

Copyright message

© AquaSPICE Consortium, 2022-2023

This deliverable contains original unpublished work except where indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both. Reproduction is authorised if the source is acknowledged.

TABLE OF CONTENTS

| | | |
|--------|--|----|
| 1. | Executive summary..... | 7 |
| 2. | Introduction..... | 8 |
| 3. | Data Management Approach | 9 |
| 3.1. | Data processing and management..... | 9 |
| 3.1.1. | Overview of Data Management Methods..... | 9 |
| 3.1.2. | AquaSPICE Data Harmonisation and Management Approach | 10 |
| 3.2. | Data Storage..... | 11 |
| 3.2.1. | Overview of Data Storage Approaches | 11 |
| 3.2.2. | AquaSPICE Storage Approach | 11 |
| 4. | AquaSPICE Consumers’ Data Requirements..... | 13 |
| 4.1. | Water Quality data..... | 15 |
| 4.2. | Process data | 16 |
| 4.3. | Economic data..... | 18 |
| 4.4. | Environmental & Energy Data | 18 |
| 4.5. | Other / auxiliary data | 19 |
| 4.6. | AquaSPICE CPS generated data..... | 21 |
| 5. | Data Models | 22 |
| 5.1. | Existing Models | 22 |
| 5.1.1. | Smart Data Models | 22 |
| 5.1.2. | SAREF Family..... | 23 |
| 5.2. | AquaSPICE Ontology | 25 |
| 5.3. | Data Harmonization | 28 |
| 6. | Conclusions and Next Steps..... | 34 |
| 7. | Appendix: Data Management Questionnaires..... | 35 |
| 7.1. | Analytics Component | 35 |
| 7.2. | Optimisation Component..... | 37 |
| 7.3. | Simulation & Modelling Component..... | 38 |
| 7.4. | LCA Component | 39 |
| 8. | References..... | 40 |

LIST OF FIGURES

| | |
|---|----|
| Figure 3-1. CRISP-DM phases..... | 9 |
| Figure 5-1. SAREF4WATR overview | 26 |
| Figure 2. AquaSPICE Core Model..... | 27 |
| Figure 5-3. Water properties model..... | 28 |
| Figure 5-4. Water flow model..... | 28 |

LIST OF TABLES

| | |
|--|----|
| Table 1. AquaSPICE Data Providers..... | 13 |
| Table 2. AquaSPICE Data Consumers..... | 14 |

ABBREVIATIONS/ACRONYMS

| | |
|---------|---|
| CRISPDM | Cross Industry Standard Process for Data Mining |
| dLCA | dynamic Life Cycle Assessment |
| DT | Digital Twin |
| EIP | Enterprise Integration Patterns |
| ETL | Extract-Transform-Load |
| ICT | Information and Communication Technologies |
| MDM | Multi-Dimensional Data Modelling |
| KDD | Knowledge Discovery Databases |
| PSM | Process Simulation and Modelling |
| RTM | Real Time Monitoring |
| SEMMA | Sample, Explore, Modify, Model, Assess |

1. Executive summary

This deliverable investigates the data requirements related to the AquaSPICE cyber physical system and its components. It describes a seamless data management & harmonisation process for effective integration of heterogeneous data sources, varying in: i) content (e.g., data from physical sensors and information systems); ii) quality (e.g., from precise well-defined data produced by static, trusted knowledge sources to dynamic data from external sources); and iii) speed of update (e.g., updated every few hours to highly dynamic, real-time, streaming data).

The deliverable also highlights key recommendations and requirements for semantically representing data to allow harmonisation and integration of heterogeneous sensor data with other data sources. State of the art storage approaches are reviewed and selected, allowing large datasets to be stored across heterogeneous computer resources and enable storage scalability. The proposed approach enables a resource abstraction layer both providing and consuming data from sensors to automation and real-time control of system resources management.

Due to the non-existence of the SynDi plant in CS#4, this CS is considered as void case and no work related to this CS is reported in this deliverable.

2. Introduction

This deliverable reports on the outcomes of task T3.2 “Smart Network with Multi-Source Data Acquisition, Harmonisation and Processing Framework”. The task investigated the data requirements related to the AquaSPICE Cyber Physical System and its constituent components. It describes a seamless data management process for effective integration of heterogeneous data sources, varying in: i) content (e.g., data from physical sensors and information systems); ii) quality (e.g., from precise well-defined data produced by static, trusted knowledge sources to dynamic data from external sources); and iii) speed of update (e.g., updated every few hours to highly dynamic, real-time, streaming data).

The deliverable describes the requirements and architecture for an underlying semantic framework to allow normalisation and integration of in situ sensor data with other data sources. The deliverable takes into account the project taxonomy & ontology, which is the outcome of Task 1.4 “Data Models, Taxonomy and Ontology for Water Use in the Process Industry”, reported in Deliverable 1.5 “Data Models, Taxonomy and Ontology Industrial Water Use.” This deliverable also reports on the state-of-the-art data storage approaches, analysing prominent choices and making recommendations for adoption and use in AquaSPICE. Data storage should allow for large datasets to be stored across heterogeneous computer resources and enable storage scalability. On top of storage, a management mechanism will allow for increased ‘in-memory’ processing and should have capabilities to deal with large amounts of streaming data. The approach should enable a resource abstraction layer both providing and consuming data from sensors to automation and real-time control of system resources management.

The research results presented herein will be used in Task 3.3 “Data Fusion Middleware”, which will implement a framework that will integrate information from different data sources and will provide contextual information to support data visualisation and analytics. Moreover, the results will be utilised in Task “4.2 “Physical System Virtualization and Knowledge Core Representation”, which will provide a full digital representation of all knowledge related to the AquaSPICE taxonomy & ontology and the process models developed in T4.1 “Dynamic Water-aware Process Modelling”, with an aim of supporting the simulation models, water efficiency data analytics and dynamic LCA assessment and problem detection services, developed in WP4 “Digital Twins and Smart Analytics”.

3. Data Management Approach

3.1. Data processing and management

3.1.1. Overview of Data Management Methods

Among its objectives, the AquaSPICE Cyber Physical System will integrate data related to water treatment and management processes. A seamless data management process for effective integration of heterogeneous data sources should cater for both static and dynamic (streaming) data and should adhere to the broad stages of big-data management. Towards the development of the AquaSPICE data processing and management pipeline of technologies, we briefly survey prominent methods.

ETL (Extract-Transform-Load) has been proposed as a method for data processing towards the creation of Data Warehouses (Simitsis, Vassiliadis & Sellis 2005); since then ETL has been utilised in various data-intensive activities, including Data and Business Analytics (Souibgui et al. 2019). ETL comprises the following four steps: (i) Extract; (ii) Clean; (iii) Transform and (iii) Load.

KDD refers to knowledge discovery from data (Piatetsky-Shapiro et al., 1996) and emphasizes data mining approaches to extract knowledge from data. KDD comprises seven steps (Shafique & Qaiser, 2014): (i) Domain Understanding; (ii) Data Selection; (iii) Pre-processing; (iv) Transformation; (v) Data Mining; (vi) Interpretation/Evaluation; (vii) Knowledge Usage.

CRISP-DM (Azevedo, et al., 2008) was introduced by Daimler-Benz, Statistical Package for the Social Sciences (SPSS) and NCR Corporation and aims to offer a holistic methodology for data analysts with an emphasis on data mining. CRISP-DM comprises six phases or stages (Figure 1, Shearer, 2000): (i) Business Understanding; (ii) Data Understanding; (iii) Data Preparation; (iv) Modelling; (v) Evaluation; and (vi) Deployment.

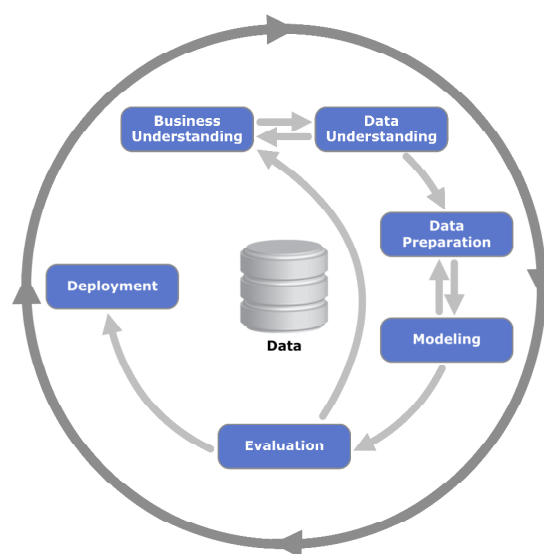


Figure 3-1. CRISP-DM phases

SEMMA (Shafique & Qaiser, 2014) was proposed by the SAS institute as a method to enable comprehension, management and evolution of data mining initiatives. It comprises five phases: (i) Sample; (ii) Explore; (iii) Modify; (iv) Model; and (v) Access.

Beyond data processing in batches, the data processing scenarios of AquaSPICE pose also the need for a streaming scenario methodology, which will be able to handle live incoming data. In such a scenario, challenges include how to solve heterogeneous sensor fusion (Kenda et al. 2019), autonomous data cleaning, potential schema changes etc. CRISP-DM has been extended to include data preparation approaches for sequence mining and to address real-time data processing scenarios (Kalgotra & Sharda 2016) such as stream analytics and signal processing in an effort to identify specific patterns in streaming data that e.g., indicate some of the outcomes such as failure of a device, etc. Moreover, CRISP-DM has been compared against SEMMA and KDD for the purpose of implementing machine learning in signal processing (Dåderman & Rosander 2018). Hence, in AquaSPICE, we will adopt the CRISP-DM methodology as the overall methodological framework for data management, adapting it to the need and the specificities of the project. A

3.1.2. AquaSPICE Data Harmonisation and Management Approach

The AquaSPICE approach to data harmonisation and management will rely on the CRISP-DM model as described above and in details in [15]. At this point, the project is in the phase of defining the data management and data harmonisation requirements with a focus on answering questions related to how data provided by the use cases need to be managed and provided as input to the AquaSPICE components. As described in this deliverable, we surveyed and analysed data requirements from AquaSPICE data consumers, i.e., CPS components, and we examined data availability by AquaSPICE data providers, i.e., use cases. The respective templates are available in the Appendix. Moreover, we identified gaps and we provide recommendations for data management.

ETL processes (data cleaning, data transformation, etc.) will be facilitated by the Water CPS Integration Bus (see section 11 in D5.1 “AquaSPICE Water CPS Services Specification”). Among the capabilities of the Integration Bus, are messaging, streaming, data transformation and orchestration. As described in D5.1, the WaterCPS Integration Bus will provide the capability to define and implement data pipelines to facilitate streaming, storage and processing of data in either real-time or batch-oriented fashion. With regards to technological choices for implementing ETL in AquaSPICE, there will be no universal technology because each Case Study has its own monitoring systems with specific technologies and infrastructure. Custom scripts, so-called adapters, are needed for each Case Study. For the development of Connectors, we will make use of open source

or/and European tools such as FIWARE IoT Agents¹ and Apache NiFi² and Kafka³, wherever possible.

3.2. Data Storage

3.2.1. Overview of Data Storage Approaches

At the heart of the AquaSPICE data management and harmonisation infrastructure are storage systems. In this section, we briefly review prominent technologies and propose appropriate ones for use in the project. Note that AquaSPICE requires technologies capable of operating with both data at rest (i.e., relational databases and NoSQL databases) as well as streaming data from the Real-Time Platform of AquaSPICE technical architecture.

There are several types of database models which provide the means of structuring the data, with most popular being the Relational Model. The relational model has been widely adopted and is performing and efficient. Moreover, database developers and administrators are familiar with this approach. Still, the relational model poses the need for well-defined schemata that define the way data are stored. This requirement for well-organised models have led database researchers and developers to explore new data storage approaches in the form of NoSQL databases. In a nutshell, NoSQL databases alleviate the need for well- data schemas and data relations. They work by supporting a loosely defined way of modelling, which provides greater flexibility and usability, although there is still the need for schema-checking in data access-time. Typical examples of NoSQL databases include MongoDB and Apache CouchDB.

3.2.2. AquaSPICE Storage Approach

Considering the project requirements and needs, a hybrid approach combining SQL and NoSQL based database will leverage the AquaSPICE data layer (see Figure 2). Based on our analysis of user requirements and in-project developer experience, the following database systems have been deemed appropriate:

- TimescaleDB⁴
- MongoDB⁵

The approach for developing the AquaSPICE data management and harmonization infrastructure will utilise open standards pertinent to data access and communication protocols. For example, the open standard nature of FIWARE NGSI offers programmers the ability to port their applications across different ‘Powered by FIWARE’ platforms and a stable framework for future development [16]. Additional functionality can be easily added to a Smart Solution simply by using additional FIWARE or third-party components for which the integration with the FIWARE Context Broker component is solved. This

¹ <https://www.fiware.org/developers/catalogue/>

² <https://nifi.apache.org>

³ <https://kafka.apache.org>

⁴ <https://www.timescale.com/>

⁵ <https://www.mongodb.com/>

integration is simplified since all components comply to the FIWARE NGSI standard interface, which eliminates vendor lock in. The component-based nature of a FIWARE based solution allows for re-architecting as the solution evolves according to business needs [16].

4. AquaSPICE Consumers' Data Requirements

This section elaborates the data that will be utilised by the AquaSPICE CPS components for the provision of their intelligent services. It focuses on understanding the data available from the AquaSPICE use cases that will feed the CPS and it aims to provide a deeper understanding of the required data models for a common representation across use cases. AquaSPICE data requirements emanate from the requirements of the Cyber Physical System as well as the data availability by the project Case Studies, see Table 1. A detailed description of the case studies datasets is provided in D8.4 “Data Management Plan”.

Table 1. AquaSPICE Data Providers

| # | Use Case | Description |
|---|------------------------|---|
| 1 | DS-CS3-WATERQUALITY | Water quality time series collected by corresponding sensors every 2 min. |
| 2 | DS-CS3-MERADES | Historic data related to conductivity, pH, Alkalinity, Calcium hardness, total hardness, Chloride, Sulphates, Temperature, Flow and Pressure |
| 3 | DS-CS3-IMPROVED | Historic Data related with (online data): TOC, Degassed cation conductivity, Sodium, Silica, Chlorine, Flow, Temperature, Pressure and Ph. Also include sampling (static data) that covers: Anions, Cations, TOC grab samples, Conductivity grab sample |
| 4 | DS-CS2-STREAM1 | Online information about pH, temperature and flowrate near to a GAC filter. |
| 5 | DS-CS2-LABDATA | Samples from laboratory experiments (water quality parameters) |
| 6 | DS-CS1A-CONTROL-SYSTEM | Online plant control system that includes information about the processes including: temperature, flow rate, conductivity, TOC, among others. |
| 7 | DS-CS1B-CONTROL-SYSTEM | Online plant control system that includes information about the processes including: temperature, flow rate, conductivity, TOC, among others. |
| 8 | DS-CS1A-IMPROVED | Historic Data related with (online data): TOC, Degassed cation conductivity, Sodium, Silica, Chlorine, Flow, Temperature, Pressure and Ph. Also include sampling (static data) that covers: Anions, Cations, TOC grab samples, Conductivity grab sample gathered in IMPROVED containers |
| 9 | DS-CS1B-IMPROVED | Historic Data related with (online data): TOC, Degassed cation conductivity, Sodium, Silica, Chlorine, Flow, Temperature, Pressure and Ph. Also include sampling (static data) that covers: Anions, Cations, TOC grab samples, Conductivity grab sample gathered in IMPROVED containers |

| | | |
|----|----------------------|--|
| 10 | DS-CS1A-MERADES | Historic data related to conductivity, pH, Alkalinity, Calcium hardness, total hardness, Chloride, Sulphates, Temperature, Flow and Pressure gathered in MERADES pilot plant |
| 11 | DS-CS1B-MERADES | Historic data related to conductivity, pH, Alkalinity, Calcium hardness, total hardness, Chloride, Sulphates, Temperature, Flow and Pressure gathered in MERADES pilot plant |
| 12 | DS-CS1B-WATERQUALITY | Water quality parameters measured in lab and monitored during laboratory experiments |
| 13 | DS-CS5-WATERQUALITY | Water quality parameters measured in lab and monitored for applicable water streams |
| 14 | DS-CS5-WATERQUALITY | Water quality parameters measured in lab and monitored for applicable water streams |

To guarantee that the AquaSPICE data management and harmonisation plan is adequate for all components of the Cyber Physical System (Table 2), Task 3.2 has been devoted to the recollection of requirements. This recollection has been carried out through several activities:

- Questionnaires: A data template was shared among CPS partners, which was used to register data from the available datasets relevant for the implementation of the specific CPS components. This process led to a record of several data parameters and their attributes which were analysed for commonalities in terms of focus and attributes. The template that was used is reported in the Appendix, whereas the analysis of the data gathered is provided next.
- CS workshops: As part of WP3, several rounds of technical workshops have been conducted for each CPS component. These workshops, where technical leaders of the CPS and partners of AquaSPICE attended, have been centered at discussing the water treatment processes to be implemented during the project and analysing the necessities with respect to the data required by the CPS components to deliver their services. (monitoring, data analytics, optimization, etc.).
- Bi-lateral meetings: Apart from the questionnaire and individual workshops, several bi-lateral meetings between WP3 members and CPS leaders have been conducted to clarify the necessities regarding data management.

Table 2. AquaSPICE Data Consumers

| # | Component | Description |
|---|---|--|
| 1 | TOOL-WP3- REALTIME- MONITORING- PLATFORM | A real-time monitoring platform design based on open-source solutions used to collect, transform, store and communicate real time measurements from field sensors, whilst guaranteeing security of communications. |
| 2 | TOOL-WP4- PROCESS- | A tool to represent the process and the interrelation of the inputs/outputs in each of the demo-cases. The tool can perform simulations of the system under what-if scenarios. |

| | | |
|---|--|--|
| | SIMULATION MODELLING TOOL | |
| 3 | TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION | Optimization methods based on AI for enabling the selection of different water schemes |
| 4 | TOOL-WP4-ANALYTICS | Descriptive and predictive analytics for anomaly detection, prediction of critical situations in industrial processes. |
| 5 | TOOL-WP5-LIFECYCLE | A visualization and exploration environment to aid-decision making. |

As a result of the activities, this section gives a detailed description of the collected data requirements. We describe the main data entities of the WaterCPS and their attributes. We identified the data entities in collaboration with the partners in charge of the different modules of the WaterCPS. We used a template with questions requesting feedback from partners about the data inputs and outputs of their modules. We grouped the responses in six main classes of data types:

- water quality data, including information mainly related to the assessment of quality characteristics of water streams.
- process data, including information mainly related to industrial treatment processes involving water streams as input, output or by-product.
- economic data, relevant to the water being used as influent to the plant.
- environmental & energy data, including energy consumption data as well as data pertaining to the environmental footprint of industrial water treatment processes.
- auxiliary and other types of data, including information relevant to the treatment process that may be used to perform modelling, simulation and optimisation of the treatment process.
- CPS generated data, that is data that are generated by CPS intelligent services.

The six classes are presented in detail in the following sections.

4.1. Water Quality data

Water Quality

Description: Water quality refers to the chemical, physical, and biological characteristics of water based on the standards of its usage. It is most frequently used by reference to a set of standards against which compliance, generally achieved through treatment of the water, can be assessed.

Relevant water stream: WaterStream

Input to: TOOL-WP3-REALTIME-MONITORING-PLATFORM

TOOL-WP4-PROCESS-SIMULATION MODELLING TOOL

TOOL-WP4-ANALYTICS, TOOL-WP5-LIFECYCLE

Properties

Total Organic Content (TOC) is a measure of the total amount of carbon in organic compounds in pure water and aqueous systems, e.g. 10 mg/l

Dynamic viscosity of the water, e.g. 0,001 Pa*s

Concentration of the solute/particle in the feed stream (e.g. Mg, Al, Ca, Cu, etc), e.g. 10 mg/l

Concentration of the solute/particle in the permeate stream, e.g. 0,1 mg/l

Chemical oxygen demand (COD), e.g. 20 mg/l

Biochemical oxygen demand (BOD) after 5 days, e.g. mg/l

Total alkalinity of the water, e.g. 100 mg/l

Conductivity of the water, e.g. 1000 μ S/cm

Water density, e.g. 997 kg/m³

Water turbidity, e.g., 2 NTU

Total dissolved solids, e.g. 100 mg/l

Total soluble solids, e.g. 10 mg/l

Concentration of phosphate ions in feed stream, e.g. 0,05 mg/l

Concentration of the nitrogen content of the ammonium ion in feed stream, e.g. 0,2 mg/l

Concentration of nitrate nitrogen in feed stream, e.g. 1,5 mg/l

Concentration of total nitrogen in feed stream, e.g. 3 mg/l

Concentration of microorganisms in the feed stream (e.g. E.coli, legionella spp, total coliforms, campylobacter etc.), e.g., <100 CFU/100 ml

Concentration of hydrogen peroxide in the feed stream, e.g. 0,1 g/l

Concentration of pollutants (e.g. chloride, sulphate, silica, sodium, zinc, iron, nickel, lead, copper, cadmium, etc.), e.g. 0,1 g/l

pH value of the solution, e.g. 6

Hardness of the solution, e.g. 12 °D

Normality of influent, e.g. g-eq/l

Normality of effluent, e.g. g-eq/l

Ion concentration in solution, e.g. mg/l

Ion concentration ion the exchange resin, e.g. mg/l

Ion equivalent fraction for the equilibrium condition ($C_e/C_o=1$), e.g. 0,4

Concentration of total phosphorus (P) in feed stream, e.g., 0,1 mg/lmg/l

Concentration of organic nitrogen and ammonia in feed stream, e.g., 0.1 mg/l

4.2. Process data

Process

Description: Process water data refer to parameters that are useful for optimising water-based industrial processes, such as heating, cooling, processing, cleaning and rinsing.

Relevant water stream: WaterStream

Input to: TOOL-WP3-REALTIME-MONITORING-PLATFOR

TOOL-WP4-PROCESS-SIMULATION MODELLING TOOL

TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION

TOOL-WP4-ANALYTICS, TOOL-WP5-LIFECYCLE

Properties

Flow rate of the stream used for backwashing, e.g. 10 m³/h

Time for backwashing, e.g. 0,001 h

Temperature of the water, e.g. 20 °C

Total operation time of a process, e.g. min

Reaction rate constant for oxidation process, e.g. 10⁷ 1/M*s

Mass of oxygen transferred in a reactor, e.g. 20 kg/d

Crossflow velocity in a membrane element, e.g. 1,5 m/s

Density of air that is added in a reactor, e.g. 1,2 kg/m³

Flow of air to a reactor, e.g. Nm³/h

Hydraulic retention time in a reactor, e.g. days

Sludge retention time in a reactor, e.g. days

Oxidation reduction potential, e.g. mV

Throughput volume when C/Co=0,5 for ion 2, e.g. 7,5 l

Throughput volume when C/Co=0,5 for ion 1, e.g. 22 l

Density of resin, e.g. 700 kg/m³

Amount of resin, e.g. 0,1 kg

Exchange capacity of the resin, e.g. 290 meq/kg

Resistance in electrodialysis unit, e.g. Ω

Kinetic parameters (e.g. specific growth rate, decay rate, hydrolysis rate, etc.) of various compounds in a unit, e.g. 1/d

Solute diffusivity, e.g. 0,000005 m²/h

Mass transfer coefficient of solute, e.g. m/h

Concentration near the membrane wall, e.g. 12 mol/m³

Osmotic pressure difference, e.g. 0,3 bar

Constant for the solute adsorption during filtration, e.g. 0,0001 1/s

Tank Capacity, 300 m³

Tank Chemical Limits, [TOC: 4 g/dL, COD: mg/L,]

WasteWater Stream Quantity (or Flow), e.g., 300 m³ (flow: 300 m³/h)

WasteWater Stream contamination factors, [TOC: 4 g/dL, COD: mg/L,]

| |
|---|
| Treatment Process Capacity Limits, e.g., 300 m ³ |
| Treatment Process Losses, e.g. 5 (%) |
| Treatment Process Efficiency, |
| Treatment Process Chemical Limits, [TOC: 4 g/dL, COD: mg/L,] |
| Applications Deadline, 21/02/2022 - 18:00 |
| Applications Water Demand, e.g. 300 m ³ |
| Applications Chemical Limits, e.g. [TOC: 4 g/dL, COD: mg/L,] |
| Applications Residual Quantity |
| Applications Residual Quality |
| WasteWater quantity flows, e.g. {"0-->1": 300, "1-->2": 200,...} |
| Fresh water intake quantity, {"0": 300, "1": 200,...} |

4.3. Economic data

| Economic data |
|---|
| <p>Description: Economic data relevant to the water being used as influent to the plant.</p> <p>Relevant water stream: WaterStream</p> <p>Input to: TOOL-WP3-REALTIME-MONITORING-PLATFOR TOOL-WP4-PROCESS-SIMULATION MODELLING TOOL TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION</p> |
| Properties |
| <p>Water tariff, Euros/m³</p> <p>Total cost of water, euros</p> <p>Discharge Quantity Cost, e.g. 300 EUROS/ m³</p> <p>Fresh Water Quantity Cost, e.g., 300 EUROS/ m³</p> <p>Treatment Process Cost, e.g., 300 EUROS/ m³</p> |

4.4. Environmental & Energy Data

| Environmental & Energy Data |
|---|
| <p>Description: Environmental data are related to the net environmental footprint of the treatment process. Energy data are linked to the energy consumption of the treatment process.</p> <p>Relevant water stream: WaterStream</p> <p>Input to: TOOL-WP3-REALTIME-MONITORING-PLATFOR TOOL-WP4-PROCESS-SIMULATION MODELLING TOOL TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION</p> |

| TOOL-WP4-ANALYTICS, TOOL-WP5-LIFECYCLE |
|--|
| Properties |
| Energy consumption per unit, e.g. 2 kWh/m ³ |
| Voltage, e.g. 0.1 V |
| Current, e.g. 1 A |
| Power, e.g. 100 W |
| Abiotic depletion, e.g. kgSbeq/m ³ |
| Fossil Fuel Depletion, e.g. MJ/m ³ |
| Global Warming, e.g. kgCO ₂ ,eq/m ³ |
| Ozone Layer Depletion, e.g. kgCFC-11eq/m ³ |
| Human Toxicity, e.g. kg1,4DCBeq/m ³ |
| Fresh Water Aquatic Ecotox, e.g. kg1,4DCBeq/m ³ |
| Marine Aquatic Ecotoxicity, e.g. kg1,4DCBeq/m ³ |
| Terrestrial ecotoxicity, e.g. kg1,4DCBeq/m ³ |
| Photochemical Oxidation, e.g. kgC ₂ H ₄ ,eq/m ³ |
| Acidification, e.g. kgSO ₂ ,eq/m ³ |
| Eutrophication, e.g. kgPO ₄ -3eq/m ³ |
| Freshwater Depletion, e.g. m ³ /m ³ |
| Treatment Process Power, e.g. 20 kW |
| Discharge Routes Chemical Limits, e.g. 300 m ³ |
| Discharge Quantity Limits, e.g. 300 m ³ |
| Fresh Water Quantity Limits, , e.g. 300 m ³ |
| WasteWater discharge quantity, e.g., {"0": 300, "1": 200,...} |

4.5. Other / auxiliary data

| WaterStream |
|---|
| Description: Refers to a particular water stream within the process plant |
| Relevant process (influent / effluent): process |
| Properties |
| Feed flow rate, e.g. 5 m ³ /h |
| Water flux, e.g. 1,3 m/s |
| Pressure of the permeate stream, e.g. 10000 Pa |
| Pressure of the feed stream, 100000 Pa |

Auxiliary data

Description: Auxiliary data are data relevant to the treatment process and may be used to perform modelling, simulation and optimisation of the treatment process

Relevant water stream: Water stream

Input to: TOOL-WP3-REALTIME-MONITORING-PLATFOR

TOOL-WP4-PROCESS-SIMULATION MODELLING TOOL

TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION

TOOL-WP4-ANALYTICS, TOOL-WP5-LIFECYCLE

Properties

Active area of the membrane element, e.g. 10 m²

Average size of the membrane pores, e.g. 0,03 μm

Length of the membrane element, e.g. 150 cm

Width of a membrane element, e.g. 85 cm

Hydrodynamic permemability of the clean membrane, e.g. 100 L/m²/h/Pa

Resistance of the clean membrane, e.g. 10¹² 1/m

Compressibility index of the cake formed during filtration, e.g. 0,6

Hydrodynamic permemability of the clean membrane, e.g. 100 L/m²/h/Pa

Specific resistance of the cake formed during filtration, e.g. 10¹⁸ m/kg/Pa⁽⁻ⁿ⁾

Solute permeability, e.g. 0,00005 m/h

Filament thickness, e.g. 0,31 mm

Feed spacer channel height, e.g. 0,6 mm

Mesh size, e.g. 3,2 mm

Spacer angle, e.g. 100 rad

Iron Chloride, e.g. 6.5 mg/l

Activated Carbon, e.g. 6.5 mg/l

Sodium Hydroxide, e.g. 6.5 mg/l

Hydrochloric Acid, e.g. 6.5 mg/l

Citric Acid, e.g. 6.5 mg/l

Sulfuric Acid, e.g. 6.5 mg/l

Sodium Bisulfide, e.g. 6.5 mg/l

Antiscalant, e.g. 6.5 mg/l

Sodium hypochlorite, e.g. 6.5 mg/l

Sludge, e.g., 100 kg/l

4.6. AquaSPICE CPS generated data

| AquaSPICE CPS generated data |
|---|
| Description: Data to be generated by the AquaSPICE CPS and its constituent components. |
| Properties |
| TOOL-WP4-WATER-EFFICIENCY-OPTIMIZATION |
| WasteWater quantity flows: the paths which the wastewater streams will follow in the system accompanied with their respective quantities, e.g. {"0-->1": 300, "1-->2": 200,...} |
| Fresh water intake quantity: the routes where fresh water will be drawn in the system accompanied with their respective quantities, e.g. {"0": 300, "1": 200,...} |
| WasteWater discharge quantity: the routes where wastewater will be discharged from the system accompanied with their respective quantities, e.g. {"0": 300, "1": 200,...} |
| TOOL-WP4-ANALYTICS |
| Total Organic Content (prediction / anomaly alert), e.g. mg/l |
| Electroconductivity (prediction / anomaly alert), e.g. $\mu\text{S}/\text{cm}$ |
| corrosion rate (prediction / anomaly alert) |
| pH (prediction / anomaly alert) |
| COD (prediction / anomaly alert) e.g., mg/l |
| H2O2 (prediction / anomaly alert) e.g., mg/l |
| N-NO3 (prediction / anomaly alert) e.g., mg/l |
| TOOL-WP5-LIFECYCLE |
| Abiotic depletion, e.g. kgSbeq/m ³ |
| Fossil Fuel Depletion, e.g. MJ/m ³ |
| Global Warming, e.g. kgCO ₂ ,eq/m ³ |
| Ozone Layer Depletion, e.g. kgCFC-11eq/m ³ |
| Human Toxicity, e.g. kg1,4DCBeq/m ³ |
| Fresh Water Aquatic Ecotox e.g. kg1,4DCBeq/m ³ |
| Marine Aquatic Ecotoxicity, e.g. kg1,4DCBeq/m ³ |
| Terrestrial ecotoxicity, e.g. kg1,4DCBeq/m ³ |
| Photochemical Oxidation, e.g. kgC ₂ H ₄ ,eq/m ³ |
| Acidification, e.g. kgSO ₂ ,eq/m ³ |
| Eutrophication, e.g. kgPO ₄ -3eq/m ³ |
| Freshwater Depletion, e.g. m ³ /m ³ |

5. Data Models

5.1. Existing Models

5.1.1. Smart Data Models

The Smart Data Models⁶ initiative compiles more than 400 data models for various application domains including water management. The models are released with an open license allowing the user the free use, free modification, and free sharing of the modifications without any other restriction. Source for the models is available at GitHub⁷. It is a collaborative effort, so it's possible to evolve the model and to suggest modifications and new data models. Data models are grouped into subjects and domains. This means, the smart data models envisions to be generated in each of the FIWARE domains (smart cities, agrifood, water, energy, environment, smart sensing, smart aeronautics, smart destination, cross-sector, smart robotics, smart health and smart manufacturing).

This Smart Data Models initiative has these principles:

- Open: Using open licensing to allow extensive use and adoption (allows free use, free modification, and free sharing for all forever).
- Community: Data models are provided by the users not created by the initiative
- Coherence: The initiative checks that the defined data model defines properties that are consistent (same name means same data type in it)
- Evolution: Allow quick evolution of data model without heavy / slow review processes
- NGSI compatible: Compatible with NGSI standard but also beyond that standard.
- Experienced: Based on real case scenarios (so data models are tested in real case scenarios before accepted)
- Adopting what is adopted: Based on widely adopted standards (including ontologies and international schemas (i.e., schema.org)
- Git based: Based on git platform and GitHub as development frontend

Considering the Smart Data Models already published, the ones that could be reused for the AquaSPICE project are:

| Smart Data Models | AQUASPICE alignment |
|------------------------------|--|
| Wastewater | Representation of wastewater treatment process and the subsequent systems and assets that generates data (e.g. tanks, simulation results, blower, etc.). |
| WaterQuality | Representation of water quality observations inside the water network. |

⁶ <https://smartdatamodels.org>

⁷ <https://github.com/smart-data-models>

| | |
|---|--|
| Waste Management | This data model represent typically involved waste management scenarios. It could serve to waste data modelling from industrial processes. |
| Alert | This data model will serve to determine alerts coming from the different devices and systems |
| Device | This data model will permit to represent the IoT devices and their corresponding nature inside the context-broker. |
| Key Performance Indicator | This data model will support the representation of the key performance indicators and their interrelationships with industrial KPIs. |
| Satellite Imagery | This data model will serve to represent EO analysis and observations across the AquaSPICE case-studies. |

5.1.2. SAREF Family

Originally, the Smart Applications REference ontology (SAREF) is published by ETSI and intended to enable interoperability between solutions from different providers and among various activity sectors on the Internet of Things (IoT), thus contributing to the development of the global digital market.

SAREF specifies recurring core concepts in smart applications, the main relationships between these concepts, and axioms to constrain the usage of these concepts and relationships. The creators of SAREF have followed a few design principles that are in line with those specified by the European Interoperability Framework (as listed in the previous section), namely: principles of reuse and alignment of concepts and relationships that are defined in existing assets, modularity to allow separation and recombination of different parts of the ontology depending on specific needs, extensibility to allow further growth of the ontology, and maintainability to facilitate the process of identifying and correcting defects, accommodate new requirements, and cope with changes in (parts of) SAREF.⁸ SAREF has been extended to a broad of domains to be established as a cross-domain interrelationship between IoT information. Indeed, SAREF includes the following extensions:

1. SAREF4ENER: extension for the Energy domain
2. SAREF4ENVI: extension for the Environment domain
3. SAREF4BLDG: extension for the Building domain
4. SAREF4CITY: extension for the Smart Cities domain
5. SAREF4INMA: extension for the Industry and Manufacturing domains
6. SAREF4AGRI: extension for the Smart Agriculture and Food Chain domains
7. SAREF4AUTO: extension for the Automotive domain (under development)

⁸Source: ETSI TS 103 264

8. SAREF4EHAW: extension for the eHealth/Ageing-well domain
9. SAREF4WEAR: extension for the Wearables domain
10. SAREF4WATR: extension for the Water domain

The continuous addition of more extensions turns SAREF into the standards family that enables better integration of semantic data from and across various vertical domains in the IoT world. For the AquaSPICE project the related reference ontologies to be considered are represented in the following table as depicted in the D1.5 entitled as “Data Models, Taxonomy and Ontology Industrial Water Use” (see Section 5.1.4 of D1.5):

| Ontology | Description |
|----------------|--|
| WAWO+ ontology | Waste Water Ontology (WaWO) (http://www.cs.upc.edu/~loliva/OntoWAWO+.owl) has been developed to model information about wastewater treatment tasks through the definition of the basic terms and relations comprising the vocabulary of the wastewater treatment area. It is the demonstration of a shared understanding of the wastewater domain that is agreed among several agents: mainly, experts in chemical, environmental and engineering. The main aim of the WaWO ontology is to design a model that: presents a vocabulary for the wastewater domain that all agents can jointly use and understand by describing the meaning of each term in an as unambiguous and precise manner as possible. |
| SAREF4WATR | This ontology is mainly focused on representing water infrastructure and the corresponding measurements. Despite this focus, the ontology considers water quality monitoring and the alignment of different water quality events with different types of waters. |
| SAREF4ENVI | SAREF4ENVI has two main aims: on the one hand, to be the basis for enabling the use of SAREF in the environment domain and, on the other hand, to exemplify how to enable interoperability between environmental devices in cooperation. |
| SAREF4INMA | SAREF4INMA focuses on extending SAREF for the industry and manufacturing domain to solve the lack of interoperability between various types of production equipment that produce items in a factory and, once outside the factory, between different organizations in the value chain to uniquely track back the produced items to the corresponding production equipment, batches, material and precise time in which they were manufactured. |
| SAREF4ENER | SAREF4ENER focuses on extending SAREF for energy management in buildings and other infrastructures. |
| WSIS | The WSIS ontology refers to model industrial symbiosis and related indicators to the different regionalities. |
| OFIS | The ontological framework for Industrial Symbiosis (IS) exploits semantic knowledge modelling and enables structural data transformation for |

| Ontology | Description |
|-------------|--|
| | identification of potential synergies between various industries and hence formation of one to one and complex symbiotic networks (Trokana et al. 2013) |
| e-Symbiosis | A new ontology framework for Industrial Symbiosis by pioneering the use of ontology engineering in the field. Semantics are used to model Industrial Symbiosis flows, to model enabling technologies and to systematise the development of a matching service (Cecelja et al. 2015). |
| MODA | A web portal devote to the categorization of models equations and simulations in relation to industrial symbiosis and process modelling. |
| CHADA | A data management portal and framework devoted to material characterization based on datasets that covers (i) sample, (ii) method, (iii) raw data and (iv) data analysis as the main component of the metadata associated to any characterization experiment (Romanos et al. 2019). |

5.2. AquaSPICE Ontology

As stated in the D1.5 “Data Models, Taxonomy and Ontology”, the AquaSPICE ontology is being constructed under the industrial representation of the IIoT systems and their connection with observations and measurements is performed using the ontological patterns represented in the SAREF and SAREF4WATR ontologies. To avoid duplication in the documentation of SAREF4WATR, specific details can be found in the reference documentation.

In a nutshell, SAREF follows the Observation and Measurement pattern for the representation of measurements. Under this pattern, specific measurements (saref:Measurements) are interlinked with specific Devices (saref:Device) to catalogue the specific digital system that performs the observation procedure. Both devices and measurements refer to a specific property (saref:Property). This property refers to the specific types of measurement (e.g., water flow, temperature, energy flows, etc.). These properties and subsequent measurements are observed in real elements or feature of interests (saref:FeatureOfInterest). These real objects normally are related to a geospatial representation (geosp:Feature) that covers points or specific areas (e.g., polygons, multi-polygons, etc.). Geospatial features are modelled using the Open Geospatial Consortium standards: <https://www.ogc.org/standards/geosparql>. Measurements are specified by a temporal specification (time:TemporalEntity) and a unit of measure (saref:UnitOfMeasure). There exists a special measurement related to the feature of interest to measure specific general indicators over a region. Figure 5-1 presents an overview of the classes and the properties included in the SAREF4WATR extension.

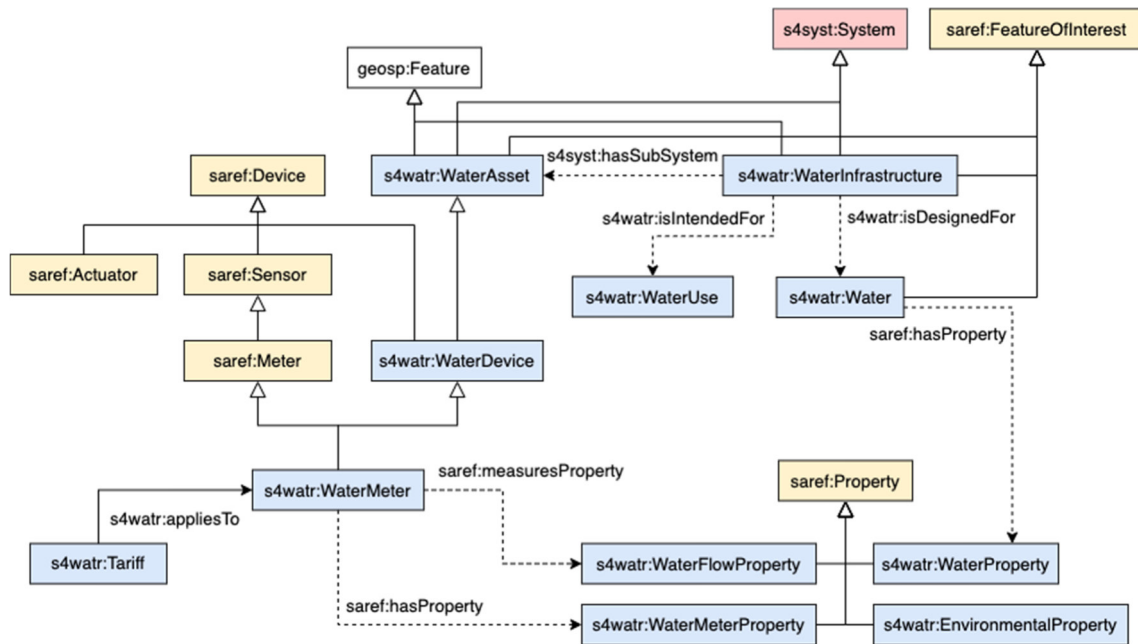


Figure 5-1. SAREF4WATR overview

Under the general structures of the SAREF and SAREF4WATR ontologies, the AquaSPICE ontology has built the interrelationships between resources and the specification of the corresponding industrial variables accordingly to several processes to be monitored and controlled under the project. This aspect has served to represent under the model the main industries, processes and technologies used. These terms have served to propose a data mechanism to identify what kind of processes and technology are applied in each of the project industries. Complementary to this aspect, we represented into the core ontology (Figure 2) the interrelationship between the process technologies and the different resources generated or consumed. This last representation permit to advance on the understanding on the quantities of resources generated and consumed by each process, industry and cluster/regionalty.

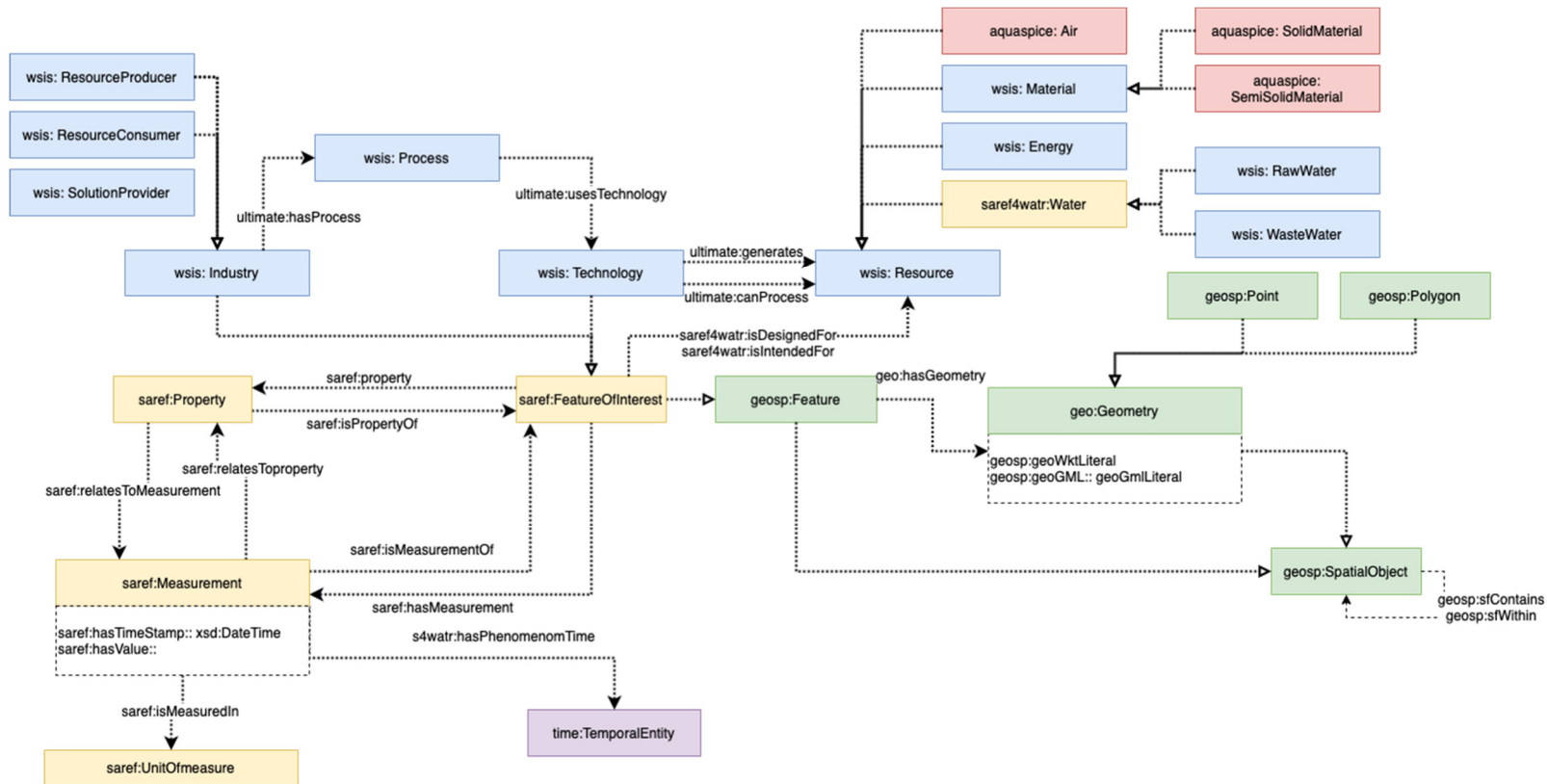


Figure 2. AquaSPICE Core Model

5.3. Data Harmonization

Considering the data requirements and the provisions available in the AquaSPICE data models, ontology and references models, namely saref and sf4watr, the next step is to harmonise between the former and the latter. The following tables present mappings between the data requirements of the AquaSPICE data consumers (described in section 4) and data models, where available. Where a mapping cannot be inferred, we indicate how the data element can be linked to reference models and submodels of sf4watr (Figure 5-3) and saref (Figure 5-4).

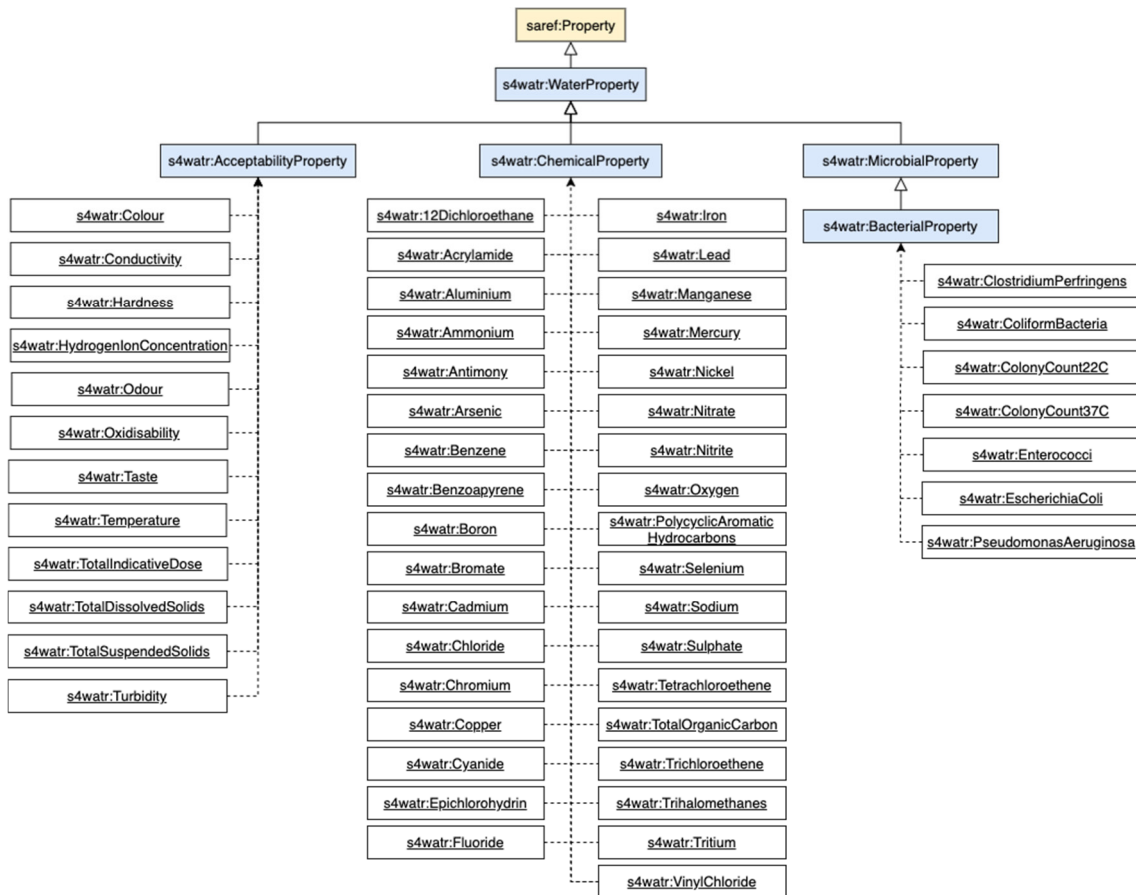


Figure 5-3. Water properties model

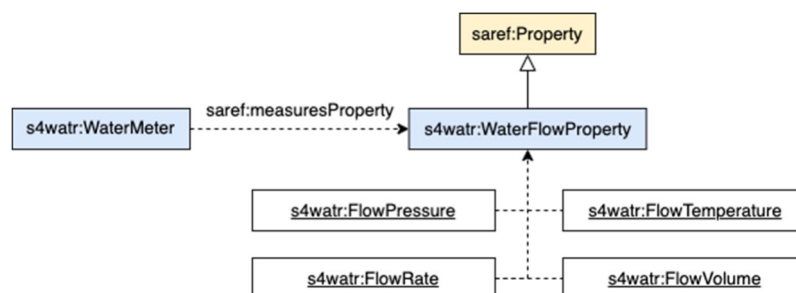


Figure 5-4. Water flow model

| Water Quality | |
|--|--|
| Properties | Data Model Mapping |
| Total Organic Content (TOC) | https://saref.etsi.org/saref4watr/TotalOrganicCarbon |
| Viscosity | NEW! Under s4watr:AcceptabilityProperty |
| Concentration of the solute/particle in the feed stream (e.g. Mg, Al, Ca, Cu, etc.) Concentration of pollutants (e.g. chloride, sulfate, silica, sodium, zinc, iron, nickel, lead, copper, cadmium, etc.) | Iron (https://saref.etsi.org/saref4watr/Iron) Lead (https://saref.etsi.org/saref4watr/Lead) Manganese https://saref.etsi.org/saref4watr/Manganese Mercury https://saref.etsi.org/saref4watr/Mercury Nickel https://saref.etsi.org/saref4watr/Nickel Copper https://saref.etsi.org/saref4watr/Copper Calcium NEW! Magnesium NEW! |
| Chemical oxygen demand (COD) | Oxygen, aka Oxygen saturation, aka Dissolved oxygen |
| Biochemical oxygen demand (BOD) | https://saref.etsi.org/saref4watr/Oxygen |
| Total alkalinity | NEW! Under s4watr:AcceptabilityProperty |
| Conductivity | Conductivity https://saref.etsi.org/saref4watr/Conductivity |
| Density | NEW! Under s4watr:AcceptabilityProperty |
| Turbidity | Turbidity https://saref.etsi.org/saref4watr/Turbidity |
| Total dissolved solids | Total dissolved solids https://saref.etsi.org/saref4watr/TotalDissolvedSolids |
| Total soluble solids | Total soluble solids https://saref.etsi.org/saref4watr/TotalSuspendedSolids |
| Concentration of phosphate ions | NEW! Under s4watr:ChemicalProperty |
| Concentration of the nitrogen | NEW! Under s4watr:ChemicalProperty |
| Concentration of microorganisms in the feed stream (e.g. E.coli, legionella spp, total coliforms, campylobacter etc.) | Bacterial property https://saref.etsi.org/saref4watr/BacterialProperty with members s4watr:ClostridiumPerfringensni, s4watr:ColonyCount37Cni, s4watr:ColonyCount22Cni, s4watr:Enterococcini, s4watr:ColiformBacteriani, s4watr:EscherichiaColini, s4watr:PseudomonasAeruginosani |
| Concentration of hydrogen peroxide | NEW! Under s4watr:ChemicalProperty |
| pH | NEW! Under s4watr:AcceptabilityProperty |
| Hardness | Hardness https://saref.etsi.org/saref4watr/Hardness |

| | |
|---|---|
| Normality | NEW! Under s4watr:AcceptabilityProperty |
| Ion concentration | NEW! Under s4watr:ChemicalProperty |
| Concentration of total phosphorus | NEW! Under s4watr:ChemicalProperty |
| Concentration of organic nitrogen and ammonia | NEW! Under s4watr:ChemicalProperty |

| Process | |
|--|--|
| Properties | Data Model Mapping |
| Flow rate of the stream used for backwashing | Flowrate: https://saref.etsi.org/saref4watr/FlowRate |
| Time for backwashing Total operation time of a process | has phenomenon time: https://saref.etsi.org/saref4watr/hasPhenomenonTime |
| Temperature of the water | Flow temperature: https://saref.etsi.org/saref4watr/FlowTemperature |
| Reaction rate constant for oxidation process | NEW! Under s4watr:TreatmentPlant |
| Mass of oxygen transferred in a reactor | NEW! Under s4watr:TreatmentPlant |
| Crossflow velocity in a membrane element | NEW! Under s4watr:TreatmentPlant |
| Density of air that is added in a reactor | NEW! Under s4watr:TreatmentPlant |
| Flow of air to a reactor | NEW! Under s4watr:TreatmentPlant |
| Hydraulic retention time in a reactor | NEW! Under s4watr:TreatmentPlant |
| Sludge retention time in a reactor | NEW! Under s4watr:TreatmentPlant |
| Oxidation reduction potential | NEW! Under s4watr:TreatmentPlant |
| Throughput volume when $C/Co=0,5$ for ion | NEW! Under s4watr:TreatmentPlant |
| Density of resin | NEW! Under s4watr:TreatmentPlant |
| Amount of resin | NEW! Under s4watr:TreatmentPlant |
| Exchange capacity of the resin | NEW! Under s4watr:TreatmentPlant |
| Resistance in electrodialysis unit | NEW! Under s4watr:TreatmentPlant |
| Kinetic parameters (e.g. specific growth rate, decay rate, hydrolysis rate, etc.) of various compounds | NEW! Under s4watr:TreatmentPlant |

| | |
|--|--|
| Solute diffusivity | NEW! Under s4watr:TreatmentPlant |
| Mass transfer coefficient of solute | NEW! Under s4watr:TreatmentPlant |
| Concentration near the membrane wall | NEW! Under s4watr:TreatmentPlant |
| Osmotic pressure difference | NEW! Under s4watr:TreatmentPlant |
| Constant for the solute adsorption during filtration | NEW! Under s4watr:TreatmentPlant |
| Tank Capacity | NEW! Under s4watr:StorageInfrastructure |
| Tank Chemical Limits | NEW! Under s4watr:StorageInfrastructure |
| WasteWater Stream Quantity | See WaterQuality |
| WasteWater Stream contamination factors | See WaterQuality |
| Treatment Process Capacity Limits | NEW! Under s4watr:TreatmentPlant |
| Treatment Process Losses | NEW! Under s4watr:TreatmentPlant |
| Treatment Process Efficiency | NEW! Under s4watr:TreatmentPlant |
| Treatment Process Chemical Limits | NEW! Under s4watr:TreatmentPlant |
| Applications Deadline | NEW! Under s4watr:TreatmentPlant |
| Applications Water Demand | NEW! Under s4watr:TreatmentPlant |
| Applications Chemical Limits | NEW! Under s4watr:TreatmentPlant |
| Applications Residual Quantity | NEW! Under s4watr:TreatmentPlant |
| Applications Residual Quality | NEW! Under s4watr:TreatmentPlant |
| WasteWater quantity flows | Saref4water WaterFlow model and properties |
| Fresh water intake quantity | Saref4water WaterFlow model and properties |

| Economic data | |
|---------------------------|--|
| Properties | Data Model Mapping |
| Water tariff | WaterTariff: https://saref.etsi.org/saref4watr/Tariff |
| Total cost of water | NEW! Under https://saref.etsi.org/saref4watr/Tariff |
| Discharge Quantity Cost | NEW! Under https://saref.etsi.org/saref4watr/Tariff |
| Fresh Water Quantity Cost | NEW! Under https://saref.etsi.org/saref4watr/Tariff |
| Treatment Process Cost | NEW! Under s4watr:TreatmentPlant |

| Environmental & Energy Data | |
|----------------------------------|---|
| Properties | Data Model Mapping |
| Energy consumption per unit | Energy: https://saref.etsi.org/core/Energy |
| Abiotic depletion | <p>NEW! The remaining data properties cannot be linked under the Environmental property:</p> <p>https://saref.etsi.org/saref4watr/EnvironmentalProperty</p> |
| Fossil Fuel Depletion | |
| Global Warming | |
| Ozone Layer Depletion | |
| Human Toxicity | |
| Fresh Water Aquatic Ecotox | |
| Marine Aquatic Ecotoxicity | |
| Terrestrial ecotoxicity | |
| Photochemical Oxidation | |
| Acidification | |
| Eutrophication | |
| Freshwater Depletion | |
| Treatment Process Power | |
| Discharge Routes Chemical Limits | |
| Discharge Quantity Limits | |
| Fresh Water Quantity Limits | |
| WasteWater discharge quantity | |

| WaterStream | |
|---------------------------------|--|
| Properties | Data Model Mapping |
| Feed flow rate | Flow rate: https://saref.etsi.org/saref4watr/FlowRate |
| Water flux | Saref4water WaterFlow model and properties |
| Pressure of the permeate stream | Flow pressure: https://saref.etsi.org/saref4watr/FlowPressure |
| Pressure of the feed stream | |

| Auxiliary data | |
|----------------|--------------------|
| Properties | Data Model Mapping |

| | |
|--|--|
| <p>Active area of the membrane element</p> <p>Average size of the membrane pores</p> <p>Length of the membrane element</p> <p>Width of a membrane element</p> <p>Hydrodynamic permemability of the clean membrane</p> <p>Resistance of the clean membrane</p> <p>Compressibility index of the cake formed during filtration</p> <p>Hydrodynamic permemability of the clean membrane</p> <p>Specific resistance of the cake formed during filtration</p> <p>Solute permeability</p> <p>Filament thickness</p> <p>Feed spacer channel height</p> <p>Mesh size</p> <p>Spacer angle</p> <p>Iron Chloride</p> <p>Activated Carbon</p> <p>Sodium Hydroxide,</p> <p>Hydrochloric Acid</p> <p>Citric Acid</p> <p>Sulfuric Acid</p> <p>Sodium Bisulfide</p> <p>Antiscalant</p> <p>Sodium hypochlorite</p> <p>Sludge</p> | <p>NEW! The auxiliary family of data can be linked under the System class of the saref for system reference model:</p> <ul style="list-style-type: none"> • System <ul style="list-style-type: none"> ◦ Water asset <ul style="list-style-type: none"> ▪ Sink asset <ul style="list-style-type: none"> ▪ Estuary ▪ Ocean ▪ River ▪ Sea ▪ Source asset <ul style="list-style-type: none"> ▪ Glacier ▪ Lagoon ▪ Lake ▪ Storage asset <ul style="list-style-type: none"> ▪ Aquifer ▪ Reservoir ▪ Tank ▪ Transport asset <ul style="list-style-type: none"> ▪ Intake ▪ Main <ul style="list-style-type: none"> ▪ Channel ▪ Pipe ▪ Maintenance hole, aka Manhole ▪ Pit ▪ Vent |
|--|--|

6. Conclusions and Next Steps

Based on the main gaps exposed under D1.5 and data requirements emanating from the WaterCPS, this deliverable harmonises the project data requirements with the corresponding semantic models under the Smart Data Models initiative and the SAREF and SAREF4WATR reference models. The approach followed uses models are open licensed and based on open standards, and they are designed to work with open-source platforms like FIWARE platform. The approach allows for transparency; it is openly available with works in progress are also visible to allow public collaboration. The approach is not defining standards but providing agile standardization to those based on actual standards. Further, the approach is technologically neutral and allows for data portability. The data models gathered in Smart Data Models count with the ability to be used with all their defined attributes, with part of them, or for particular purposes be extended with their own attributes for every user. It prevents any kind of vendor lock-in. Finally, the approach allows for normalisation and integration of in situ sensor data with other data sources, describing and interlinking the resource producers and consumers.

Based on the approach described herein, Task 3.3 “Data Fusion Middleware” will implement a framework that will integrate information from different data sources and will provide contextual information to support data visualisation and analytics. Moreover, Task 4.2 will develop a full digital representation of all knowledge, relating also to the T1.4 taxonomy & ontology and the process models developed in T4.1, with an aim of supporting the simulation models, water efficiency data analytics and dynamic LCA assessment and problem detection services.

The resulting technology will:

- Handle information structured in arbitrary ways. Basic data types such as character arrays, numbers, booleans nested in arbitrary complex structures.
- Allow normalisation and integration of in situ sensor data with other data sources in the network.
- Facilitate the integration of real time data coming from different sources (sensors, weather prediction services, etc...).
- Include an access control mechanism that will allow to define a context-based set of rules to determine consent provision aspects and preferences to be applied in either automatic or manual manner.

7. Appendix: Data Management Questionnaires

7.1. Analytics Component

| | | | | |
|--|---|--|-----------------|--------------------|
| | Name: | Analytics | | |
| | Responsible partner: | ICCS | | |
| | Short description: | This component will implement descriptive and predictive analytics | | |
| | | | | |
| | | | | |
| | Name | Description | Category | Coming from |
| Input 1 | e.g. TOC | e.g. water quality metric | quality | CS1B, CS1T, |
| Total organic content | BFW TOC | total organic content of the boiler feed water | quality | CS1B |
| Electrical conductivity | BFW EC | boiler feed water electrical conductivity | process | CS1B |
| Ferrite | BFW Fe _{total} | boiler feed water ferrite | quality | CS1B |
| Copper | BFW Cu | boiler feed water copper | quality | CS1B |
| Silicon dioxide | BFW SiO ₂ | boiler feed water silicon dioxide | quality | CS1B |
| Calcium/Magnesium | BFW Ca/Mg | boiler feed water calcium/magnesium | quality | CS1B |
| pH | BFW pH[-] | boiler feed water ph | quality | CS1B |
| Calcium | Ca | calcium in feed | quality | CS1B |
| Total organic content | TOC[mg/L] | Total organic content | quality | CS1B |
| Calcium | CTMU Ca | CTMU Calcium | quality | CS1B |
| Chemical oxygen demand | CTMU COD | CTMU Chemical oxygen demand | quality | CS1B |
| Corrosion rate | Corrosion rate | | quality | CS1B |
| | CoC vs. Thickening | | | CS1B |
| Chemical oxygen demand | CT West COD | CT west Chemical oxygen demand | | CS1B |
| Chemical oxygen demand | CT P1 COD[mg/L] | Chemical oxygen demand | | CS1B |
| Chemical oxygen demand | CT P2 COD[mg/L] | Chemical oxygen demand | | CS1B |
| Zinc | CT West Zn[mg/L] | Zinc | | CS1B |
| Zinc | CT P1 Zn[mg/L] | Zinc | | CS1B |
| Zinc | CT P2 Zn[mg/L] | Zinc | | CS1B |
| phosphorus | CT West P _{tot} [mg/L] | phosphorus | | CS1B |
| phosphorus | CT P1 P _{tot} [mg/L] | phosphorus | | CS1B |
| phosphorus | CT P2 P _{tot} [mg/L] | phosphorus | | CS1B |
| Adsorbable Organically bound halogens | CT West AOX[µg/L] | Adsorbable Organically bound halogens | | CS1B |
| Adsorbable Organically bound halogens | CT P1 AOX[µg/L] | Adsorbable Organically bound halogens | | CS1B |
| Adsorbable Organically bound halogens | CT P2 AOX[µg/L] | Adsorbable Organically bound halogens | | CS1B |
| pH | Witznitz pH[-] | pH | | CS1B |
| Total organic content | Witznitz TOC[mg/L] | Total organic content | | CS1B |
| Chemical oxygen demand | Witznitz COD[mg/L] | Chemical oxygen demand | | CS1B |
| Turbidity | Witznitz Turbidity[NFU] | Turbidity | | CS1B |
| Electrical conductivity | Witznitz EC | Electrical conductivity | | CS1B |
| | any other variables given in Witznitz | | | |
| | | | | |
| | any variable that is given for Pulgar I CTBD and possibly also CTMU | | | |
| | any variable that is given for Pulgar I CTBD and possibly also CTMU | | | |
| | | | | |
| pH | TV1 - pH | Influent to the pilot system | | CS2 |
| Chemical oxygen demand | TV2 - COD | Influent to the AOP column | | CS2 |
| Peroxide | TV3 - H ₂ O ₂ | Influent to the AOP column | | CS2 |
| pH | TV4 - pH | Influent to the AOP column | | CS2 |
| Chemical oxygen demand | TV5 - COD | Influent to the anoxic aerobic MBR | | CS2 |
| Nitrate | TV6 - N-NO ₃ | Influent to the anoxic aerobic MBR | | CS2 |
| Peroxide | TV7 - H ₂ O ₂ | Influent to the anoxic aerobic MBR | | CS2 |
| Chemical oxygen demand | TV8 - COD | Effluent from the anoxic aerobic MBR | | CS2 |

| | | | |
|------------------------|--------------------------------|---|------|
| Nitrate | TV9 - N-NO3 | Effluent from the anoxic aerobic MBR | CS2 |
| pH | pH | Influent to the equalization tank | CS2 |
| Chemical oxygen demand | COD | Influent to the equalization tank | CS2 |
| HRT | HRT | equalization tank | CS2 |
| Peroxide | H2O2 | Influent to the equalization tank | CS2 |
| pH | pH | Influent to the equalization tank | CS2 |
| Soda caustic | Soda Caustic Dosed | equalization tank | CS2 |
| Chemical oxygen demand | COD | Influent to the AOP column | CS2 |
| AOP | AOP Column Adsorption | AOP Column | CS2 |
| Peroxide | H2O2 | Influent to the AOP column | CS2 |
| AOP fouling | Fouling of AOP column | AOP column | CS2 |
| Nitrate | NO3 | Influent to the Equalization Tank | CS2 |
| Chemical oxygen demand | COD | Influent to the MBR system | CS2 |
| HRT | HRT | Aerobic tank of MBR | CS2 |
| Temperature | Temperature | MBR | CS2 |
| Dissolved oxygen | Dissolved oxygen concentration | Aerobic tank of MBR | CS2 |
| pH | pH | Aerobic tank of MBR | CS2 |
| total suspended solids | TSS (total suspended solids) | MBR | CS2 |
| Nitrate | NO3 | Influent to the MBR system | CS2 |
| Glycerole | Glycerole addition | Anoxic Aerobic MBR | CS2 |
| pH | pH | Anoxic Aerobic MBR | CS2 |
| pH | pH | Influent to the MBR system | CS2 |
| Peroxide | H2O2 | Influent to the MBR system | CS2 |
| Dissolved oxygen | Dissolved oxygen | Anoxic compartment of MBR | CS2 |
| | | | |
| Output 1 | e.g. anomaly alert | e.g., description of anomaly in data stream | |
| DOW | anomaly alert | BFW TOC[mg/L] | CS1B |
| | anomaly alert | BFW EC[μS/cm] | CS1B |
| | anomaly alert | corrosion rate | CS1B |
| | prediction | BFW TOC[mg/L] | CS1B |
| | prediction | BFW EC[μS/cm] | CS1B |
| | prediction | corrosion rate | CS1B |
| | | | CS1B |
| Solvay | anomaly alert | TV1 - pH | CS2 |
| | anomaly alert | TV2 - COD | CS2 |
| | anomaly alert | TV3 - H2O2 | CS2 |
| | anomaly alert | TV4 - pH | CS2 |
| | anomaly alert | TV5 - COD | CS2 |
| | anomaly alert | TV6 - N-NO3 | CS2 |
| | anomaly alert | TV7 - H2O2 | CS2 |
| | anomaly alert | TV8 - COD | CS2 |
| | anomaly alert | TV9 - N-NO3 | CS2 |
| | prediction | TV1 - pH | CS2 |
| | prediction | TV2 - COD | CS2 |
| | prediction | TV3 - H2O2 | CS2 |
| | prediction | TV4 - pH | CS2 |
| | prediction | TV5 - COD | CS2 |
| | prediction | TV6 - N-NO3 | CS2 |
| | prediction | TV7 - H2O2 | CS2 |
| | prediction | TV8 - COD | CS2 |
| | prediction | TV9 - N-NO3 | CS2 |

7.2. Optimisation Component

| | | Name: | Optimisation | | | | | | | |
|----------|--------------|--------------------|---|-----------------|--------------|----------------------|----------------------|--|--|--|
| | | Responsible: | AUEB | | | | | | | |
| | | Short description: | | | | | | | | |
| | Component | Name | Description | Data type | Category | Example | Format | Communication frequency | | |
| Input 10 | Treatment Pr | Treatment Pr | The power of each treatment process | float | energy | 20 kW | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 5 | Discharge R | Discharge R | Chemical limits of discharging wastewater | vector of float | environmenta | vector of float | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 6 | Discharge R | Discharge Qu | Limits regarding the maximum quantity which can be discharged on each discharge route (e.g. as set by a contract) | float | environmenta | 300 m³ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 8 | Fresh Water | Fresh Water | Limits regarding the maximum quantity which can be drawn (e.g. as set by a contract or due to system limitations) | float | environmenta | 300 m³ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Output 3 | WasteWater | WasteWater | The routes where wastewater will be discharged from the system accompanied with their respective quantities | dict of floats | environmenta | {"0": 300, "1": 300} | field of a JSON file | LIVE | | |
| Input 7 | Discharge R | Discharge Qu | Cost of discharging some wastewater quantity | float | financial | 300 EUROS/ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 9 | Fresh Water | Fresh Water | Cost of fresh water intake quantity | float | financial | 300 EUROS/ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 13 | Treatment Pr | Treatment Pr | The cost of each Treatment Process | float | financial | 300 EUROS/ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 21 | Other | Any short term | Special constraints provided to the optimization solver by the plant managers | undefined | other | undefined | field of a JSON file | LIVE when opt algorithm is activated | | |
| Input 1 | Tank | Tank Capacit | The capacity of each Tank included in the system | float | process | 300 m³ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 2 | Tank | Tank Chemica | Chemical limits of each tank as specified by the manufacturer | vector of float | process | [TOC: 4 g/dL | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 3 | WasteWater | WasteWater | Quantity (or flow) of waste water on each stream | float | process | 300 m³ (flow: | field of a JSON file | LIVE when opt algorithm is activated | | |
| Input 4 | WasteWater | WasteWater | Chemical density of contamination factors (pollutants) on each wastewater stream | vector of float | process | [TOC: 4 g/dL | field of a JSON file | LIVE when opt algorithm is activated | | |
| Input 11 | Treatment Pr | Treatment Pr | The capacity of each Treatment Process included in the system | float | process | 300 m³ | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 12 | Treatment Pr | Treatment Pr | The percentage of water losses occurring on each Treatment Process | float | process | 5 (%) | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 14 | Treatment Pr | Treatment Pr | The efficiency of each Treatment Process, as defined by modeling | Unknown (ba | process | Unknown | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 15 | Treatment Pr | Treatment Pr | Chemical limits for the safe operation of each Treatment Process | vector of float | process | [TOC: 4 g/dL | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 16 | Applications | Applications | The time up to when some water quantity must be available at the application location | datetime/string | process | 21/02/2022 - | field of a JSON file | LIVE when opt algorithm is activated | | |
| Input 17 | Applications | Applications | Water Quantity needed at the application location | float | process | 300 m³ | field of a JSON file | LIVE when opt algorithm is activated | | |
| Input 18 | Applications | Applications | Chemical limits for the safe use of the water quantity sent to each application | vector of float | process | [TOC: 4 g/dL | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 19 | Applications | Applications | Water Quantity leaving the applications site - based on modeling | Unknown (ba | process | Unknown | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Input 20 | Applications | Applications | Water Quality leaving the applications site - based on modeling | Unknown (ba | process | Unknown | field of a JSON file | Either hard constrained OFFLINE, or LIVE when opt algorithm is activated | | |
| Output 1 | WasteWater | WasteWater | The paths which the wastewater streams will follow in the system accompanied with their respective quantities | dict of floats | process | {"0->1": 300, | field of a JSON file | LIVE | | |
| Output 2 | Fresh Water | Fresh Water | The routes where fresh water will be drawn in the system accompanied with their respective quantities | dict of floats | process | {"0": 300, "1": 300} | field of a JSON file | LIVE | | |

7.3. Simulation & Modelling Component

| Name | Description | Category | Coming from |
|-----------------------|--|------------|--------------------------------|
| | Name: Simulation and Modeling | | |
| | Responsible: TUC | | |
| | Short descrip: Information needed for modeling and simulation | | |
| | Name: Description | | |
| solute flux | J _s solute flux | contextual | CS1T, CS1B, VS3 |
| cost of water | C _{water} total cost of water | economic | CS2 |
| water tariff | WT water tariff | economic | CS2 |
| current | i current | energy | CS1T |
| energy cons | E energy consumption per unit | energy | CS1T |
| power | P power | energy | CS1T |
| voltage | E _v voltage | energy | CS1T |
| Area | A _m active area of the membrane element | other | all CS |
| Compressibi | n compressibility index of the cake formed during filtration | other | CS1T, CS1B |
| filament thic | dF filament thickness | other | CS1T, CS1B, VS3 |
| height | h feed spacer channel height | other | CS1T, CS1B, VS3 |
| Length | L length of the membrane element | other | all CS |
| Membrane re | R _m resistance of the clean membrane | other | all CS |
| mesh size | IM mesh size | other | CS1T, CS1B, VS3 |
| permeability | L _p hydrodynamic permeability of the clean membrane | other | all CS |
| Pore size | p average size of the membrane pores | other | all CS |
| solute perme | B solute permeability | other | all CS |
| spacer angle | θ spacer angle | other | CS1T, CS1B, VS3 |
| Specific cak | α specific resistance of the cake formed during filtration | other | CS1T, CS1B |
| width | b width of a membrane element | other | all CS |
| Adsorption c | bi constant for the solute adsorption during filtration | process | CS1T, CS1B |
| air density | ρ _{air} density of air that is added in a reactor | process | CS2, CS5 |
| air flow | Q(A) flow of air to a reactor | process | CS2, CS5 |
| backwash d | t(back) time for backwashing | process | CS1T, CS1B |
| Backwash fl | Q _{back} flow rate of the stream used for backwashing | process | CS1T, CS1B |
| Concentratio | C _m Concentration near the membrane wall | process | CS1T, CS1B, VS3 |
| crossflow ve | v crossflow velocity in a membrane element | process | all CS |
| diffusivity | D _i solute diffusivity | process | CS1T, CS1B, VS3 |
| EDR resistar | R resistance in electro dialysis unit | process | CS1T |
| effluent conc | C _{eff} solute concentration in the effluent stream (e.g. TDS, TSS, BOD5, COD, P, Fe, etc.) | process | all CS |
| exchange ca | EC exchange capacity of the resin | process | CS1T, CS1B, CS2, CS3, CS5 |
| Feed flow | Q _{in} feed flow rate | process | all CS |
| Feed pressu | P _f pressure of the feed stream | process | all CS |
| hydraulic ret | HRT hydraulic retention time in a reactor | process | CS2, CS5 |
| kinetic parar | k kinetic parameters (e.g. specific growth rate, decay rate, hydrolysis rate, etc.) of various compounds in a unit | process | CS2, CS5 |
| mass transfe | k(m) mass transfer coefficient of solute | process | CS1T, CS1B, VS3 |
| operation tim | t total operation time of a process | process | all CS |
| osmotic pres | ΔΠ osmotic pressure difference | process | CS1T, CS1B, VS3 |
| oxidation rec | ORP oxidation reduction potential | process | CS2 |
| oxygen mas: | Q(O ₂) mass of oxygen transferred in a reactor | process | CS2, CS5 |
| Permeate co | C _p concentration of the solute/particle in the permeate stream | process | all CS |
| Permeate flo | Q _p permeate flow rate | process | all CS |
| Permeate pri | P _p pressure of the permeate stream | process | all CS |
| pressure diff | ΔP pressure difference between feed and pressure stream | process | all CS |
| reaction rate | k _i reaction rate constant for oxidation process | process | CS1B, CS2, CS3, CS5 |
| recovery | S recovery | process | all CS |
| resin density | ρ _{resin} density of resin | process | CS1T, CS1B, CS2, CS3, CS5 |
| resin mass | W _{resin} amount of resin | process | CS1T, CS1B, CS2, CS3, CS5 |
| resin volume | V _{resin} volume of resin | process | CS1T, CS1B, CS2, CS3, CS5 |
| Retentate co | C _r solute concentration in the retentate stream | process | all CS |
| Retentate flo | Q _r retentate flow rate | process | all CS |
| sludge reten | SRT sludge retention time in a reactor | process | CS2, CS5 |
| solute reject | R _i rejection of a specific solute | process | all CS |
| Temperature | T temperature of the water | process | all CS |
| throughput v | V1 throughput volume when C/Co=0.5 for ion 1 | process | CS1T, CS1B, CS2, CS3, CS5 |
| throughput v | V2 throughput volume when C/Co=0.5 for ion 2 | process | CS1T, CS1B, CS2, CS3, CS5 |
| volume betw | V _{ions} throughput volume between ions breakthrough curves at C/Co=0.5 | process | CS1T, CS1B, CS2, CS3, CS5 |
| water flux | J _w water flux | process | all CS |
| | | process | CS1B, CS1T, CS2, CS3, CS5 |
| Alkalinity | Alk total alkalinity of the water | quality | all CS |
| Biochemical | BOD5 biochemical oxygen demand after 5 days | quality | CS1T, CS1B, CS2, CS5, CS6 |
| Chemical ox: | COD chemical oxygen demand | quality | CS1T, CS1B, CS2, CS3, CS5, CS6 |
| concentratio | C(H ₂ O ₂) concentration of hydrogen peroxide in the feed stream | quality | CS1B, CS2, CS3, CS5 |
| concentratio | C _{input} concentration of microorganisms in the feed stream (e.g. E.coli, legionella spp, total coliforms, campylobacter etc.) | quality | all CS |
| conductivity | k conductivity of the water | quality | CS1T, CS1B, VS3 |
| density | ρ water density | quality | all CS |
| effluent norm | N _{eff} normality of effluent | quality | CS1T |
| Feed concen | C _f concentration of the solute/particle in the feed stream (e.g. Mg, Al, Ca, Cu, etc) | quality | all CS |
| feed concne | C _{in} concentration of pollutants (e.g. chloride, sulfate, silica, sodium, zinc, iron, etc.) | quality | all CS |
| feed normali | N _f normality of influent | quality | CS1T |
| hardness | Hardness hardness of the solution | quality | CS1T, CS2, CS3, CS5 |
| ion concentr: | C _s ion concentration in solution | quality | CS1T, CS1B, CS2, CS3, CS5 |
| ion concentr: | C _{resin} ion concentration ion the exchange resin | quality | CS1T, CS1B, CS2, CS3, CS5 |
| ion equival: | X _{ion} ion equivalent fraction for the equilibrium condition (Ce/Co=1) | quality | CS1T, CS1B, CS2, CS3, CS5 |
| normality di | ΔN difference in normality between feed and effluent | quality | CS1T |
| nitrate nitrog | N-NO ₃ - concentration of nitrate nitrogen in feed stream | quality | all CS |
| nitrogen con | N-NH ₄ ⁺ concentration of the nitrogen content of the ammonium ion in feed stream | quality | CS1T, CS1B, CS2, CS5, CS6 |
| Permeate co | C _p concentration of the solute/particle in the permeate stream | quality | all CS |
| pH | pH pH value of the solution | quality | all CS |
| phosphate io | PO ₄ -3 concentration of phosphate ions in feed stream | quality | CS1T, CS1B, CS2, CS5, CS6 |
| selectivity c | K _{ions} selectivity coefficient | quality | CS1T, CS1B, CS2, CS3, CS5 |
| Total dissolv | TDS Total dissolved solids | quality | CS1T, CS1B, CS2, CS5, CS6 |
| total Kjeldah | TKN concentration of organic nitrogen and ammonia in feed stream | quality | CS1T, CS1B, CS2, CS5, CS6 |
| Total nitroge | TN concentration of total nitrogen in feed stream | quality | CS1T, CS1B, CS2, CS5, CS6 |
| Total organic | TOC total organic content of the water | quality | all CS |
| Total phosph | TP concentration of total phosphorus (P) in feed stream | quality | CS1T, CS1B, CS2, CS5, CS6 |
| Total soluble | TSS Total soluble solids | quality | CS1T, CS1B, CS2, CS5, CS6 |
| turbidity | Turb water turbidity | quality | CS1T, CS1B, CS5, CS6 |
| Viscosity | μ dynamic viscosity of the water | quality | all CS |

8. References

- [1] Azevedo, A., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADS-DM.
- [2] Cecelja, F., N. Raafat, N. Trokanas, S. Innes, M. Smith, A. Yang, Y. Zorgios, A. Korkofygas και A. Kokossis, «e-Symbiosis: technology-enabled support for Industrial Symbiosis targeting Small and Medium Enterprises and innovation,» *Journal of Cleaner Production*, τόμ. 98, pp. 336-352, 2015
- [3] Dåderman, A., & Rosander, S. (2018). Evaluating frameworks for implementing machine learning in signal processing: A comparative study of CRISP-DM, semma and kdd. (Dissertation). Retrieved from <http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-235408>
- [4] Kalgotra, P., & Sharda, R. (2016). Progression analysis of signals: Extending CRISP-DM to stream analytics. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 2880-2885). IEEE.
- [5] Kenda, K., Kažič, B., Novak, E., & Mladenčić, D. (2019). Streaming data fusion for the internet of things. *Sensors*, 19(8), 1955.
- [6] Piatetsky-Shapiro, G., Brachman, R. J., Khabaza, T., Kloesgen, W., & Simoudis, E. (1996, August). An overview of issues in developing industrial data mining and knowledge discovery applications. In *KDD* (Vol. 96, pp. 89-95).
- [7] Piatetsky, G. (2014). CRISP-DM, still the top methodology for analytics, data mining, or data science projects. *KDD News*.
- [8] Romanos, N., Kalogerini, M., Koumoulos, E. P., Morozinis, A. K., Sebastiani, M., & Charitidis, C. (2019). Innovative Data Management in advanced characterization: Implications for materials design. *Materials Today Communications*, 20, 100541.
- [9] Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- [10] Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing*, 5(4), 13-22.
- [11] Shafique, U., & Qaiser, H. (2014). A comparative study of data mining process models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12(1), 217-222.
- [12] Simitsis, A., Vassiliadis, P., & Sellis, T. (2005, April). Optimizing ETL processes in data warehouses. In 21st International Conference on Data Engineering (ICDE'05) (pp. 564-575). IEEE.
- [13] Souibgui, M., Atigui, F., Zammali, S., Cherfi, S., & Yahia, S. B. (2019). Data quality in ETL process: A preliminary study. *Procedia Computer Science*, 159, 676-687.
- [14] Trokanas, N., T. Raafat, F. Cecelja και A. Kokossis, «OFIS – Ontological Framework for Industrial Symbiosis,» *Computer Aided Chemical Engineering*, vol. 32, pp. 523-528, 2013.
- [15] D3.2 - Transport and Spatial Data Warehouse Technical Design, Public Deliverable of Harmony H2020 project, available at: https://harmony-h2020.eu/wp-content/uploads/2020/09/D3.2-TSDW-Design_v1.pdf
- [16] FIWARE: The Open Source Platform of Choice for Building Smart Energy Solutions, available at: <https://www.fiware.org/wp-content/uploads/2018/10/Smart-Energy-Brochure-FIWARE-Web-1-2.pdf>