ASTRAL

*A*ll Atlantic Ocean **S**ustainable, Profi**T**able and **R**esilient **A**quacu**L**ture

GA no **863034**

Research and Innovation Action (RIA)

Start date: 1st September 2020. End date: 30th August 2024

D3.4 Technical specifications of the AIDAP platform

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 863034

Document administration

Deliverable number	3.4
Deliverable name	Technical specifications of AIDAP platform
WP number	3
Deliverable due date	M24 (31/8/2022)
Submission date	31/8/2022
Dissemination levels	Public
Lead beneficiary	BIOCEANOR
Responsible	Charlotte Dupont
Contributors	EGM (Ahmed Abid, Benoit Orihuela)
Internal reviewers	Dominique Durand (NORCE), Marie Smith (CSIR), Elisa Ravagnan (NORCE)

Document history and changes

Version	Date	Author	Description
1.0	27/04/2022	Charlotte Dupont	Create Table Of Content and first indication of content
1.1	01/08/2022	Charlotte Dupont, Benoit Orihuela, Ahmed Abid	First complete draft
1.2	30/08/2022	Charlotte Dupont	Final version, after revision from reviewers

Evidence of accomplishment

Report

Table of contents

1 Summary ... 5
2 Introduction ... 6
3 The AIDAP platform ... 7
4 Architecture of the platform ... 8
4.1 Preliminary work ... 8
4.2 Global architecture ... 8
4.3 Data ingest services ... 11
4.3.1 Context of the platform ... 12
4.3.2 SSO service ... 12
4.3.3 Stellio Context broker ... 13
4.3.4 MQTT Bridge ... 14
4.3.5 SenML to NGSI-LD: A use-case example ... 15
4.4 AI prediction services ... 18
4.5 Visualization services ... 19
5 Conclusion ... 21

Table of acronyms

AI	Artificial Intelligence
AIDAP	Artificial Intelligence Data Analytics Platform
API	Application Programming Interface
HTTP	HyperText Transfert Protocol
IMTA	Integrated Multi Trophic Aquaculture
MQTT	Message Queuing Telemetry Transport
NGSI-LD	Next Generation Service Interface - Linked Data
SenML	Sensor Measurement List

1 Summary

This document is part of technology development efforts of the Horizon 2020 All Atlantic Ocean Sustainable, Profitable and Resilient Aquaculture (ASTRAL) project. The deliverable describes the functional and technical architecture of the Artificial Intelligence Data Analytics Platform (AIDAP).

The goal of this platform is to give access to Artificial Intelligence (AI) analytics to farmers, and especially integrated multi-trophic aquaculture (IMTA) farmers. It provides a user-friendly interface to visualize all the important features (for example status of water quality, biomass, health, etc…) of one farm in one place, including AI advice for a sustainable and efficient production.

This document describes the main technical components of the platform, for data ingestion, AI prediction and visualizations, while giving some context to its usage and outcomes.

Data ingestion is a critical part of the platform, which allows to centralize and standardize the data coming from different sources. An important work has been done to propose a data standard dedicated to aquaculture and to build an efficient way to collect various kinds of data.

AI prediction is the core of the intelligence of the platform. While the data modelling is not done in the platform but offline, in another process, the platform offers the technical support to run the predictions in real-time as soon as the data are collected. It requires powerful AI support tools and architecture.

Data visualisation is the front end for the end users. It must be easy to use and must respond to their specific needs while being technically strong enough to support real time data and AI generated advice display. It requires a light but powerful tool to reveal the true potential of AIDAP dedicated to aquaculture users.

2 Introduction

The aquaculture industry is faced with the massive challenge of feeding a growing human population with an increasing demand for seafood, while remaining both sustainable and profitable. Sustainability in the aquaculture sector is particularly important considering overfishing issues and the global depletion of fish resources. Aquaculture is developing new codes of practice to achieve production sustainability.

Over the last few decades there have been major technological advances in smart sensors, data analytics and artificial intelligence, which have been utilised in increasing the productivity and profitability across land and aquatic cultivation systems. Agriculture started transforming into precision agriculture in the 1980s increasing both efficiency and sustainability, while precision aquaculture only started during the late 2010s. The number of sensors deployed in aquaculture farms has increased dramatically in recent years, and databases with several years of historical data are becoming available. However, the sheer volume of data and number of different data sources can be overwhelming, and most farmers do not know what to do with this amount of data. In IMTA specifically, multiple types of production are performed together, and success is based on a delicate equilibrium between the different species. Although it is essential to gather information on the different production processes, analysing these data efficiently to give smart insight on the production status, the production evolution, and how to improve the yield while also monitoring the surrounding environment, can be challenging.

The true value and potential of sensors and their data lie in the appropriate analysis of the data sources to produce relationships, analytics, and predictive capabilities that can help inform decision-making. The combination of multiple sources of high frequency data are particularly well-suited to machine learning (ML) algorithms. The definition of Machine Learning is the use and development of computer systems that can learn and adapt without following explicit instructions, by using algorithms and statistical models to analyse and draw inferences from patterns in data[1]. Therefore, Machine Learning algorithms can work only with a relatively big amount of data, the more the better.

AI and relevant data analytics are important and powerful tools which could unleash a huge potential for better and informed decision making in every step of the production process, to help achieving

[1] Oxford Dictionnary

increased productivity and sustainability. For example, farmers could be able to perform precision feeding with reduced feed loss or detect in advance a particular event like a harmful algal bloom.

However, access to this kind of resource is challenging for farmers. This cutting-edge technology is often expensive, and most farmers must generally invest in one technology or one type of AI analytics, if any. We believe that the true potential of AI analytics lies in the combination of all types of data sources and analytics in one place. Having access to one tool that combines all the above will truly unleash the potential for aquaculture sustainability, especially for IMTA. This is the goal of ASTRAL's AIDAP platform.

3 The AIDAP platform

The Artificial Intelligence Data Analytics Platform (AIDAP) is a platform that aims to give farmers and researchers access to a global set of data analytics tools, dedicated to aquaculture.

The goal of this platform is to facilitate the access to analytics and AI tools for aquaculture combined with an aggregator system of any type of data used in aquaculture. The types of users of the AIDAP platform targeted are:

- **Farmers and aquaculture stakeholders** will be able to visualize all the different data from the IMTA farm (sensors, cameras, weather forecasts, etc.) and to visualize the results of the analytics on the platform (when to feed, what to expect in the next hours, is there an HAB event happening, etc.)
- **Researchers** will be able to feed the platform with their own data and use the machinery of the AI algorithms to see the outcomes on their data and improve the models while extending the database.

For that, the AIDAP platform has two interfaces: an analytics interface (to ingest data and run the machine learning algorithms) and a visualization interface (to visualize data and AI outputs).

The main outputs of the AIDAP platform are ultra-specific advice for the users, specific to their farm, their production type, and their data. The analytics interface, maintained and improved by researchers and developers, will constantly bring more precision to the advice given to the farmer.

4 Architecture of the platform

4.1 Preliminary work

To build the AIDAP and determine the desired architecture, a preliminary study was conducted to define the requirements. We interrogated the members of consortium and asked them to fill a table with all the expected functionality of the platform. It provided a list of the different types of data that would be handled in the platform (images, timeseries, simulations models, etc..), the different users that would need access (farmers, students, regulators, etc…), the different interfaces that would be necessary to visualize data and results (user friendly interface with graphs, internal access to resource code, etc…), and the different usages of the platform (to get advice on production, to check sensors value, to apply a new AI model, etc…).

This list was then used to build use case scenarios to verify that all requirements would be covered in each scenario and constructed the architecture for AIDAP accordingly.

Example Scenario: I am a farmer, I want to visualize early warnings and proxy variables concerning IMTA environmental conditions

Requirement for AIDAP:

- Data input: environmental sensors, weather sensors & forecasts
- Data processing: real-time data processing, AI modelling
- Visualisation interface: real-time data visualisation, alerts

4.2 Global architecture

From a functional point of view, the global architecture of the platform is composed of 2 main interfaces: the analytics interface and the visualization interface (see figure 1):

- **The analytics interface**, hidden to the end-users, handles the data ingestion, data pre-processing and the AI-modelling features.
- **The visualization interface** is dedicated to the end-users and handles the real-time visualization of data coming from sensors or from the AI-modelling features.

Technically, the AIDAP platform is composed of three main components:

- **The Data Ingest Service:** the servers and software that handle the ingestion of data from the different sources, the storage of the data, and provide the Application Programming Interface (API) to serve the data when requested (from other services). An API is a necessary protocol for the interoperability of data between services (internally or externally, when sharing data safely is required).
- **The AI Prediction Service:** the servers and the software that handle the tasks of storing/managing the different versions of the AI Models, run the AI Models on the data and store prediction results in the Data Ingest Service
- **The Visualization Service:** the servers and software that handle the Frontend visualisation interface that will be used by the platform users, the data requesting from the Data Ingest Service and the management of the alerts that can be configured by the users

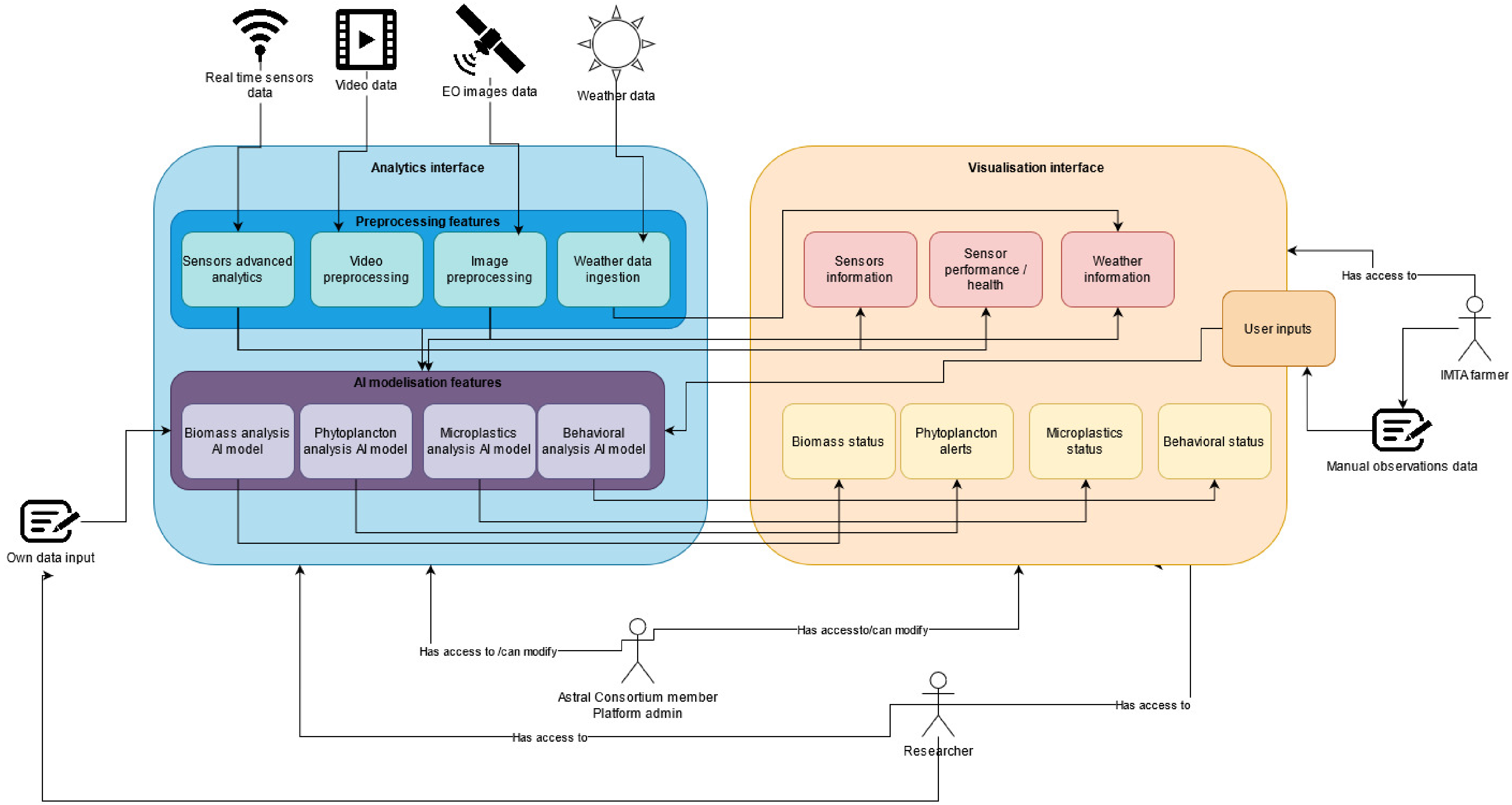

Figure 1 – Global architecture of AIDAP platform

4.3 Data ingest services

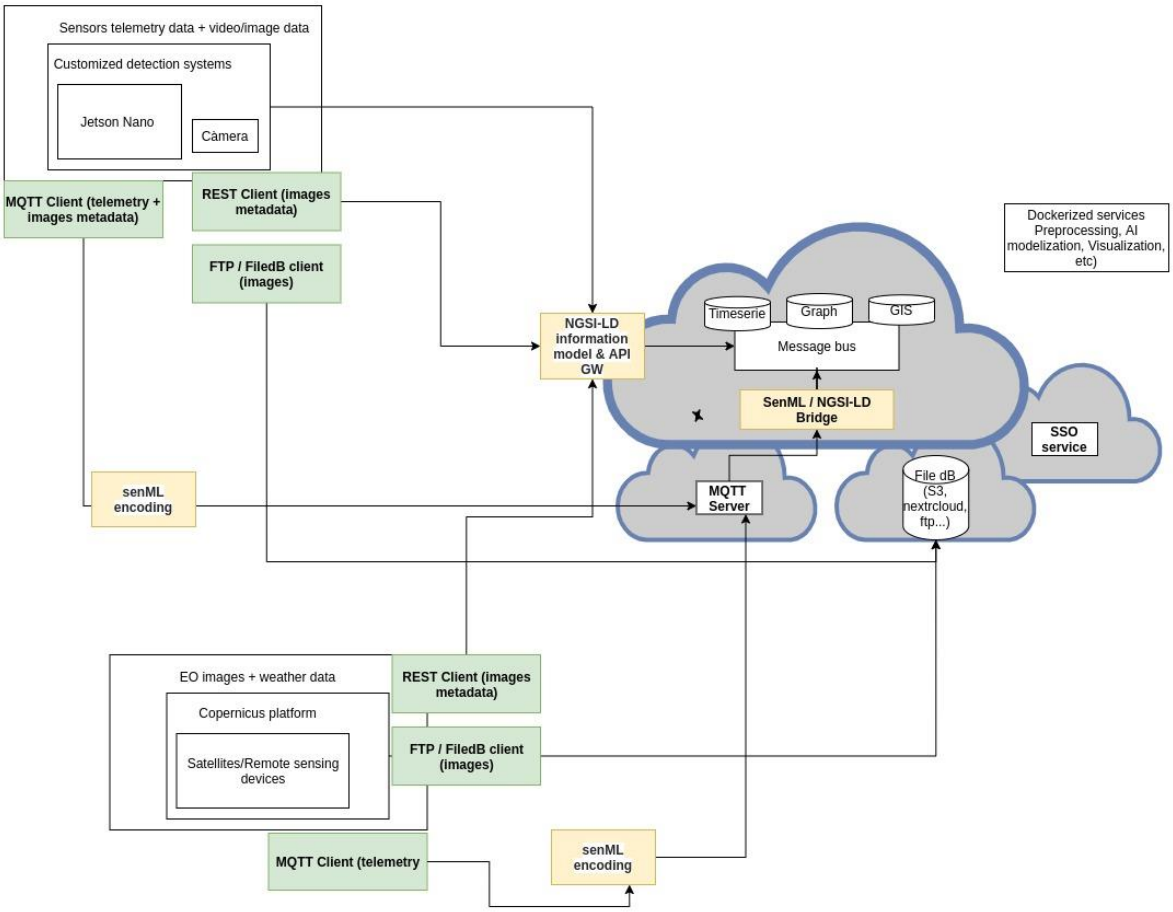

Figure 2 – Architecture of the data ingest services

The Data Ingest Service gathers, validates, transforms, and stores data coming from the different sensors, external third-party sources or manually uploaded by users. This can typically be done through:

- HTTP API: a software intermediary that allows two applications to talk to each other
- MQTT: a publish-subscribe, machine to machine network protocol
- Manual uploading of multimedia or raw data

The data can be either **Telemetry data** (sensors measurements, for example temperature = 21°C) that is encoded in **SenML,** or **Multimedia data** (for example an image of shrimps in a tank) that is handled by the **Object Storage Service**.

This service is implemented using:

- Keycloak, SSO service based on the open-source Keycloak solution
- Stellio, the FIWARE context broker developed by EasyGlobalMarket

- MQTT Bridge, a lightweight component developed by EGM, responsible for the gathering and transformation of sensor telemetry data

4.3.1 Context of the platform

The proposed platform is based on the European open-source ecosystem FIWARE which uses the NGSI-LD specification produced by ETSI:

- The FIWARE community, led by the eponymous foundation, relies on 10 years of R&D to provide contextualized data management solutions. A catalog of open-source software bricks (called "generic enablers") is made available to a community of more than 800 supplier or technology user members. Among these members can be cited the Red Hat company [2]or the Open and Agile Smart Cities (OASC) association[3]
- The NGSI-LD open specification is defined by the ISG-CIM group of the European standardization organization ETSI. The specification notably includes the definition of a REST API associated with a "cross-domain" data model allowing the exchange of heterogeneous business data between different application domains. This model is based on the principle of "linked data", derived from the semantic web, which allows the entities stored in the platform to be linked together. These two points make it possible to promote the interoperability and openness of the platform. This standard is recommended within several initiatives of the European Commission, the latest being Living-in.Eu (The European way of digital transformation in cities and communities), supporting the deployment of projects such as urban data platforms, the creation of digital twins of cities and territories, etc.

4.3.2 SSO service

The SSO service is provided by the open source Keycloak solution, developed, and maintained by RedHat. It is a mature service, very feature rich, and supported by an active community.

Keycloak implements full support for OpenID Connect and OAuth2 standards. It offers a complete administration interface allowing it to be configured very finely and regularly, and in particular:

- Graphic customization of all pages presented to users
- Configuration of security rules: expiration of inactive sessions, validity period of OAuth2 tokens, email verification at account creation time, etc.
- Activation of additional security functions, such as "two-factor authentication (2FA)" or "one-time password (OTP)"

[2] https://www.redhat.com/

[3] https://oascities.org/

- Traceability of access to applications, active sessions, etc.
- Security protections (manual revocation of tokens in the event of an identified risk, detection of brute force attacks, etc.)

Each service of the AIDAP platform is integrated with this service.

4.3.3 Stellio Context broker

Stellio is an NGSI-LD compliant context broker developed by EGM. It is a FIWARE Generic Enabler and thus can be integrated as part of any “Powered by FIWARE” platform.

In the scope of the ASTRAL project, Stellio is mainly used to store two kinds of data, and thus is making use of two different, specialized databases:

- TimescaleDB: a PostgreSQL extension designed to handle timeseries data. It is specifically suited to store telemetry data coming from sensors
- MinIO: an object storage database capable of handling large volumes of (binary or other) data. It is used to store raw and multimedia data coming from partners experiments and measures (e.g., results of spectrometry measures).

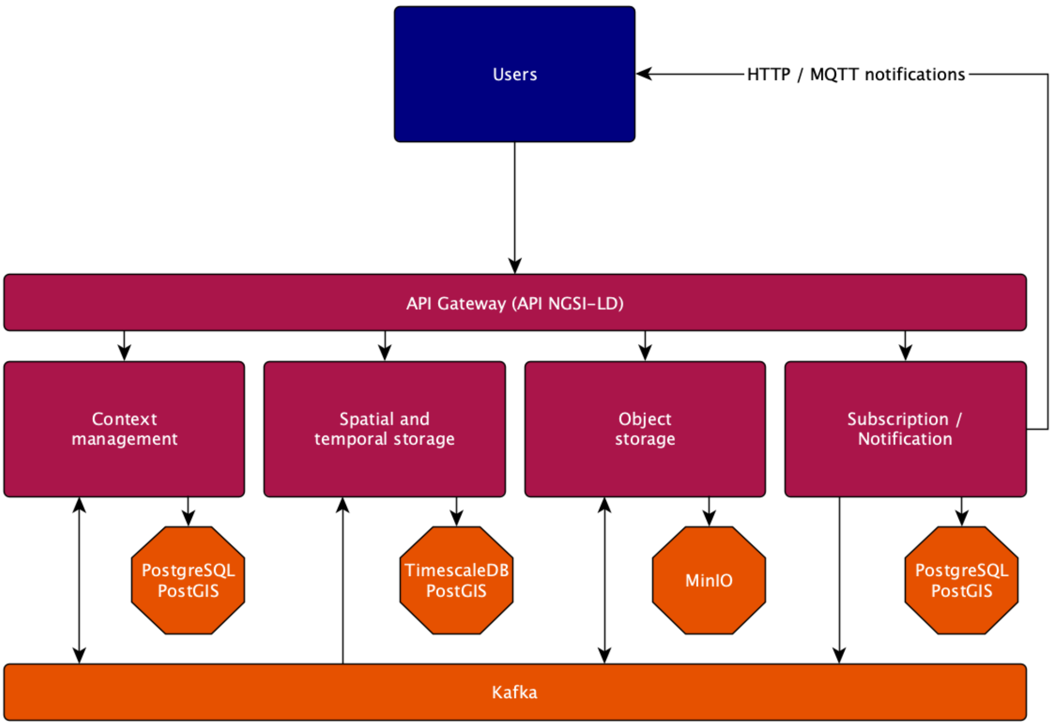

Figure 3 - Stellio architecture

To achieve these objectives, it exposes some endpoints, all suited to a specific use-case:

- Trusted components, like the MQTT Bridge, are allowed to send transformed telemetry data to the internal message bus (based on Kafka) used by Stellio
- Other external actors can use the NGSI-LD API exposed by the context broker. For instance, AI prediction services use this API to get data and publish the prediction results into the context broker.

4.3.4 MQTT Bridge

The MQTT Bridge consumes data in SenML format[4] then transforms and injects NGSI-LD events into the Stellio context broker.

The MQTT Bridge is composed of 2 main levels:

- SenML data acquisition level, in which the bridge will subscribe to specific topics of an MQTT broker and then will generate NGSI-LD events onto the context broker.
- Provisioning level, in which the bridge will subscribe to any events about entities of type mapping on the context broker.

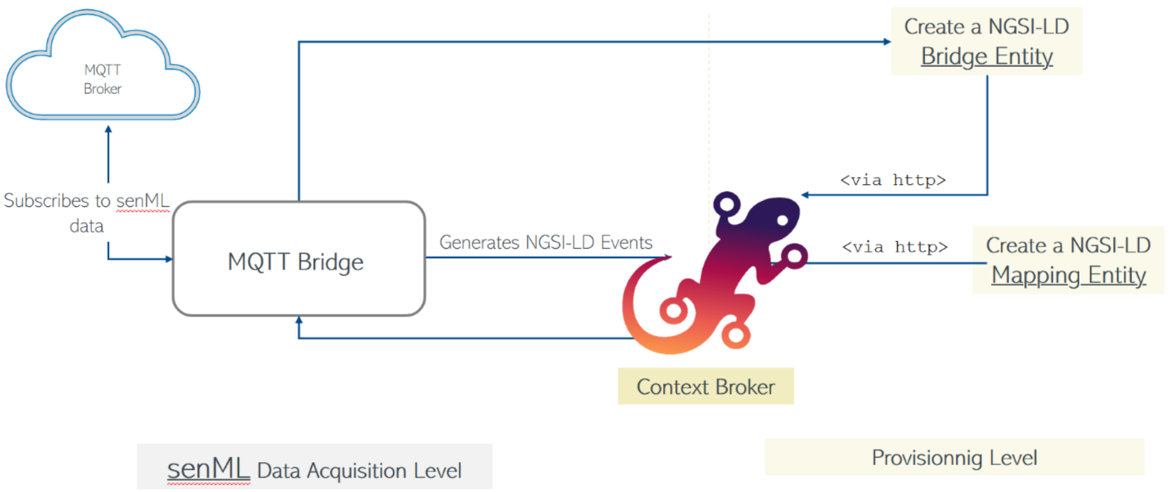

Figure 4 MQTT Bridge main architecture

[4] https://datatracker.ietf.org/doc/html/rfc8428

4.3.5 SenML to NGSI-LD: A use-case example

In this section we will describe generic use-case that consists of measuring water quality parameters in fish tanks.

4.3.5.1 NGSI-LD data model

The NGSI-LD data model of this use case is based on entities of types "Fish_Containment" used to model fish tanks, cages or units and "Device" used to model the installed devices in the fish tanks.

All water quality parameters (total dissolved solids, conductivity, dissolved Oxygen, redox, pH, temperature, and turbidity) are modelled as properties of the Fish Containment entity.

The installed device is composed of a set of probes that may be replaced for technical reason (a maintenance or an upgrade for a new model). For this purpose, the main device and probes are modelled as NGSI-LD entities. The main device contains the battery voltage property. The probes entities contain the serial number property. The main installed probes on the device are

- Probe DO: for measuring dissolved oxygen and temperature,
- Probe pH: for measuring pH, redox and temperature,
- Probe CND: for measuring conductivity, salinity, total dissolved solids and temperature,
- Probe Turb: for measuring Turbidity and temperature

The main components of the NGSI-LD data model are depicted in the figure below:

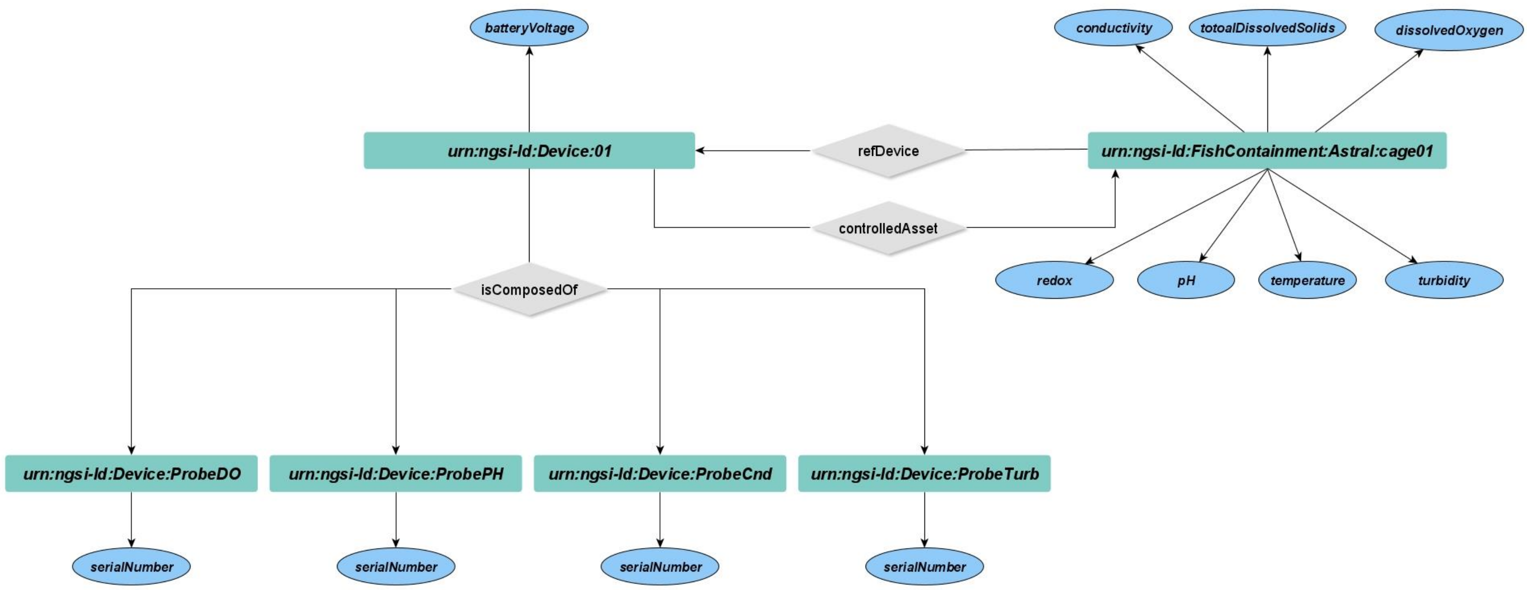

Figure 5 - NGSI-Ld data model

4.3.5.2 SenML to NGSI-LD

According to the previous sections the deployed device in the fish tank is providing senML messages via mqtt. An example of senML message received by the MQTT bridge is depicted in the figure below:


```
[{"bt":1659413318,"bn":"urn:dev:01"},{"n":"batteryVoltage","u":"V","v":4.26}
,{"n":"waterTemperatureDO","u":"cel","v":26.74},{"n":"DissolvedOxygen","u":"
%","v":58.49},{"n":"DissolvedOxygenPPM","u":"ppm","v":4.67},{"n":"DissolvedO
xygenMGL","u":"mg/l","v":4.67},{"n":"SN_DO","vs":"SN-PODOC-
2116"},{"n":"waterTemperaturePH","u":"cel","v":26.46},{"n":"PH","u":"Q30","v
":7.76},{"n":"Redox","u":"2Z","v":290.83},{"n":"SN_PH","vs":"SN-PPHRB-
6073"},{"n":"waterTemperatureCnd","u":"cel","v":26.50},{"n":"Conductivity","
u":"µs/cm","v":12.89},{"n":"Salinity","u":"g/kg","v":0.01},{"n":"TDS","u":"p
pm","v":6.51},{"n":"SN_CND","vs":"SN-PC4EB-
6200"},{"n":"waterTemperatureTurb","u":"cel","v":26.50},{"n":"TurbidityNTU",
"u":"NTU","v":27.01},{"n":"TurbidityFNU","u":"FNU","v":27.01},{"n":"Turbidit
yMGL","u":"mg/l","v":44.22},{"n":"SN_Turb","vs":"SN-PNEPB-4416"}]
```


To understand the received senML message the MQTT bridge needs some details about the destinations of received values. For example, the device `urn:dev:01` will update values of entities `urn:ngsi-ld:FishContainment:astral:01` and `urn:ngsi-ld:Device:01` according to the name of the attribute. For the `batteryVoltage` the target entity will be `urn:ngsi-ld:Device:01` and the target property is `batteryVoltage`. `For the waterTemperatureDO` the target entity is `urn:ngsi-ld:FishContainment:` and the property is `temperature`. All these details are stored in the Stellio context broker. The MQTT bridge will provision his devices according to the existing mapping entities on Stellio.

On the following, an example of the mapping entity for the device `urn:dev:01,` for the senML attributes `batteryVoltage` and `waterTemperatureDO`


```
{
   "id": "urn:ngsi-ld:Mapping:357862090230121",
   "type": "Mapping",
   "hasBridge": {
      "type": "Relationship",
      "object": "urn:ngsi-ld:Bridge:MqttBridge01"
   },
   "deviceId": {
```

```
      "type": "Property",
      "value": "urn:dev:357862090230121"
   },
   "context": {
      "type": "Property",
      "value": "https://raw.githubusercontent.com/easy-global-market/ngsild-api-data-models/master/astral/jsonld-contexts/astralCompound.jsonld"
   },
   "attributesToMap": [
      {
         "type": "Property",
         "value": "batteryVoltage",
         "attribute": {
            "type": "Property",
            "value": "batteryVoltage"
         },
         "attributeValueType": {
            "type": "Property",
            "value": "float"
         },
         "entityType": {
            "type": "Property",
            "value": "Device"
         },
         "hasTargetEntity": {
            "type": "Relationship",
            "object": "urn:ngsi-ld:Device:357862090230121"
         },
         "datasetId": "urn:ngsi-ld:Dataset:01"
      },
      {
         "type": "Property",
         "value": "waterTemperatureDO",
         "attribute": {
            "type": "Property",
            "value": "temperature"
         },
         "attributeValueType": {
            "type": "Property",
            "value": "float"
         },
         "entityType": {
            "type": "Property",
            "value": "FishContainment"
         },
         "hasTargetEntity": {
            "type": "Relationship",
            "object": "urn:ngsi-ld:FishContainment:astral:cageCSIR"
```

```
        },
        "attributeDatasetId": {
           "type": "Property",
           "value": "urn:ngsi-ld:Dataset:DO"
        },
        "datasetId": "urn:ngsi-ld:Dataset:02"
     }   ],
   "@context": [
      "https://raw.githubusercontent.com/easy-global-market/ngsild-api-data-
models/master/mapping/jsonld-contexts/mapping-compound.jsonld"
   ]
}
```


4.4 AI prediction services

The AI prediction services will be used to run the developed algorithms in ASTRAL project. For example, the model for biomass estimation of shrimps has been trained offline, with a given dataset of shrimps' images and videos in different conditions. After training and testing, the model will be uploaded on AIDAP platform to run online with new incoming data of shrimps.

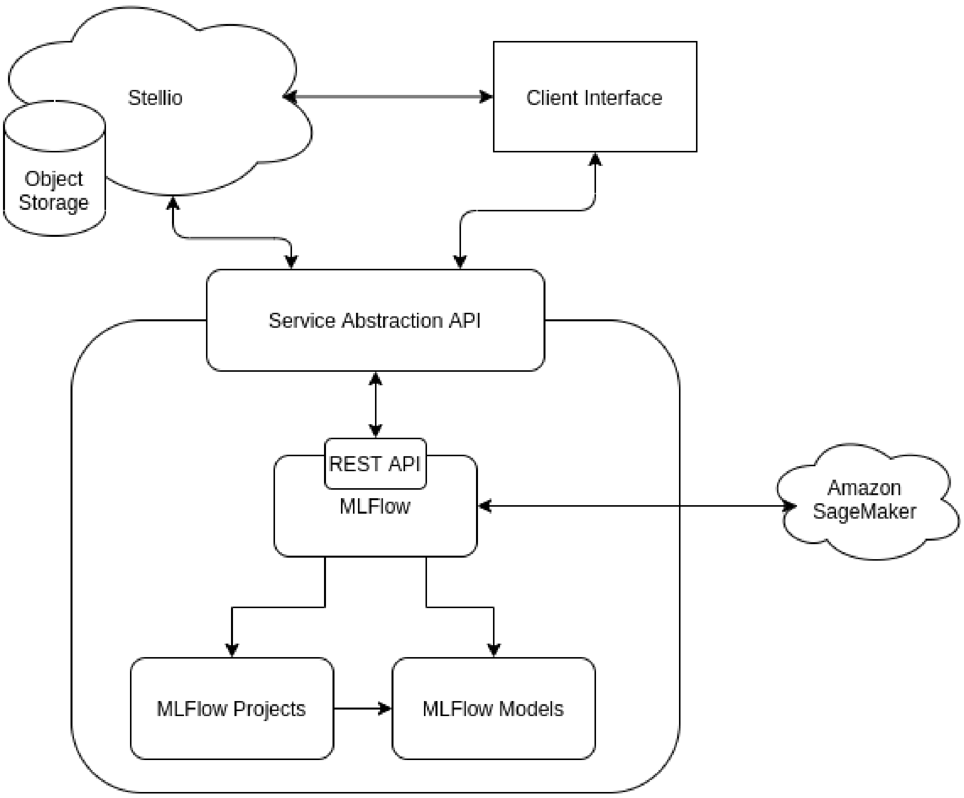

Figure 6 - AI prediction services architecture

The AI Prediction Service has the following features:

- Management of the AI Models in different Projects, and tracking of the versions of the models
- Interfacing with the Data Ingest Service for getting the data to process for the prediction, and saving the prediction result back to the service
- Launching the AI Models in a GPU machine on AWS cloud (AWS SageMaker)

This service is implemented using MLFlow as the manager of the AI models and the solution to launch the models on AWS, and custom python scripts that are implementing the interface between MLFlow and Stellio.

The AI Prediction Service is mainly using MLFlow as the solution to manage the AI Models and using AWS SageMaker to do the predictions. The second part is a Python abstraction layer that manages the connection between MLFlow and Stellio for getting the data and saving the prediction results.

4.5 Visualization services

The Visualization Service is the Frontend platform that allows to create visualisation interface for the end users. For example, the low cost IoT kits deployed in the IMTA labs in ASTRAL projects (set of sensors sending data through cellular communication or radio network) are sending many parameters data to the AIDAP and the visualisation service will allow the end users to display them on their own dashboard.

The visualization service is based on Grafana, an open-source software [5]specialized in this field. It is mature, very widespread, extremely flexible, thanks to a very complete plugin system which allows it to connect to very diverse data sources.

It natively offers the following main features:

- The provision by default of all the display widgets: simple text, statistics, gauges, tables, graphs, interactive maps, images with overlay, heat maps, etc.
- The possibility of finely configuring all graphic representations (colours, legends, reference lines, etc.)
- An organization of visualizations into dashboards and folders
- Complete graphic customization (logo, colours, etc.)

[5] https://grafana.com/

- Configuration and logging of automatic alerts (on value thresholds, absence of data, advanced calculations, etc.) that can be sent to a wide choice of channels
- Display of real-time or historical data
- User and group management, which allows to manage access rights to the various dashboards (administration, editing, reading), as well as an official extension allowing authentication to be based on an OAuth2 provider

The dashboards get the data from the Data Ingest Service and displays it in the appropriate widgets to the user.

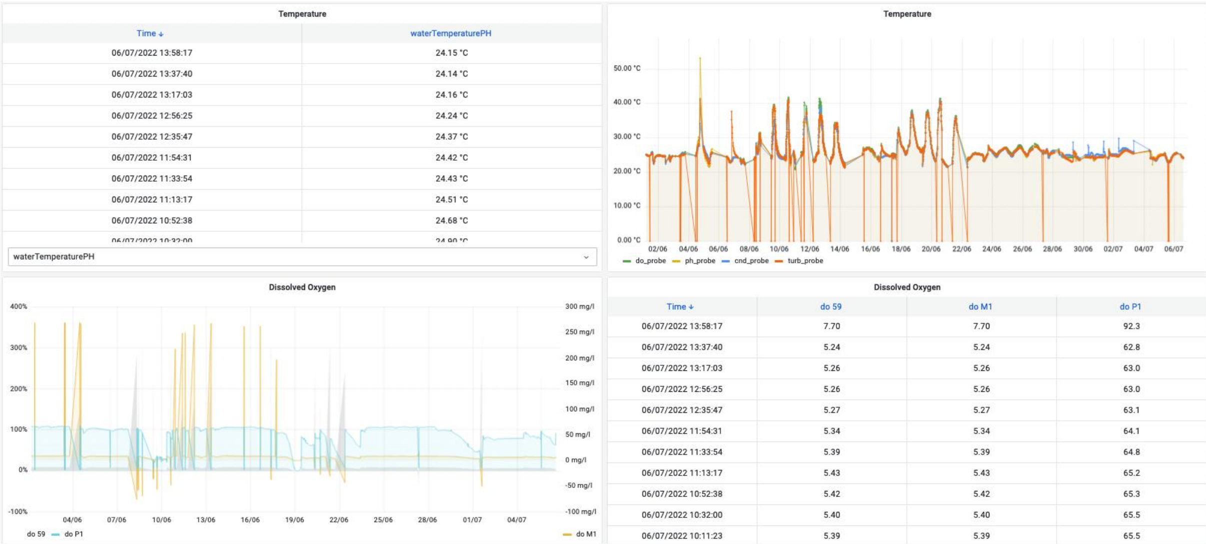

Figure 7 : Dashboard exposing real time data from Low-cost IoT Kit at CSIR

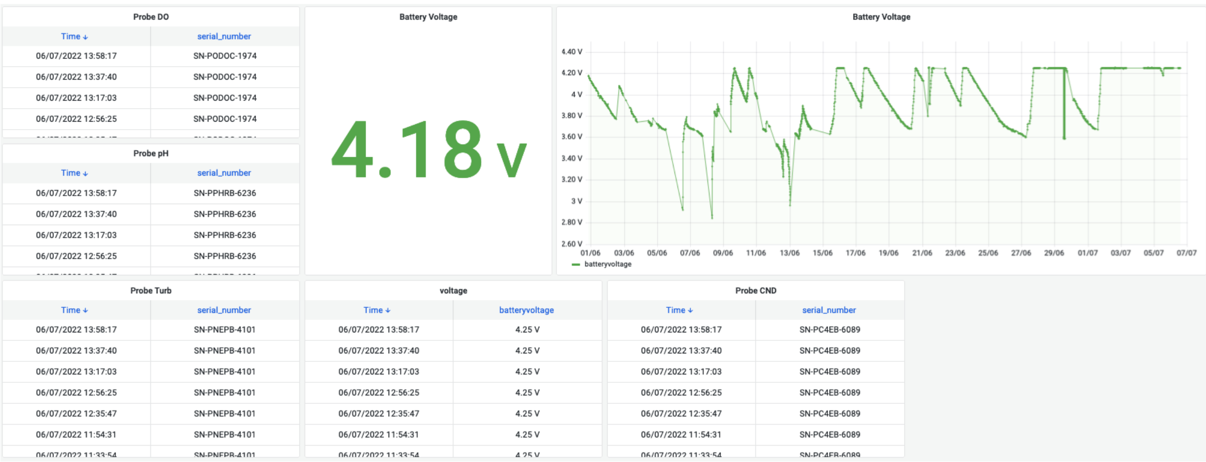

Figure 8 : Technical monitoring of a CSIR device

Some first alerts have been configured and setup. In the next steps, they will be extended to allow a very fine-grained monitoring of the data ingested.

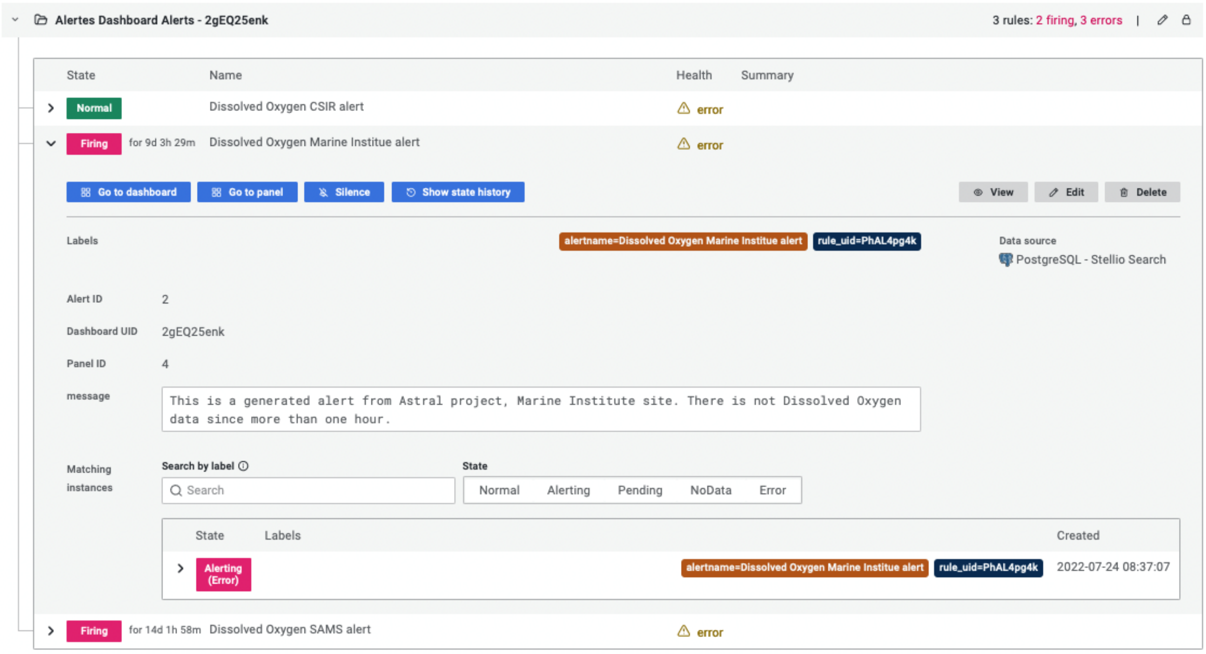

Figure 9 : Example of a monitoring alert

5 Conclusion

The AIDAP platform has been developed taking into consideration the requirements for aquaculture and IMTA especially. Using available bricks of technology, developed in several context (agriculture, aquaculture), the AIDAP platform efficiently combine their utilities and generate a powerful tool helping in the aquaculture operations and environmental monitoring.

The platform uses powerful data ingestion processes for quick data processing and quick responses. This is mandatory to be able to give real-time recommendations and take appropriate action. It also uses the best cutting-edge AI tools for easy maintenance of the developed algorithms in the project. ASTRAL is developing AI tools that AIDAP will easily made accessible to users. The visualisation of the data is intuitive and flexible to many uses, thus it can be adapted to many different IMTA farms and production type. ASTRAL has also deployed several low-cost IoT Kits and each deployment has its own specific dashboard based on their needs.

The foreseen next steps of the development of AIDAP are to receive feedbacks from end-users, to understand specific needs in the use of AIDAP. The feedbacks will be extremely useful to improve the platform and to make it the best possible tool for a real-life aquaculture IMTA farm production.