

AQUARIUS initial Data Management Plan template for TA Call applications

September 2026

Project q-fDOM: AQUARIUS initial Data Management Template - to be completed and submitted as part of the TA Call application

Project answers in italics

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

Project q-fDOM will collect three types of data

- 1) fDOM sensor responses in different laboratory conditions (temperature, water DOM content), several sensor types will be used.*
- 2) equations calculated from the abovementioned experiments and calibration data,*
- 3) describing the sensor correction functions and their interoperability*
- 4) fDOM sensor data in situ, when integrated in ferrybox systems.*

Objectives are to

- 1) Share knowledge on the use, calibration and data-correction procedures for fDOM sensors among partnerships, leading towards development of pan-European best practices in fDOM sensor calibration and validation*
- 2) Comprehensive experimental determination of fDOM temperature dependency for various sensors across the different water types and the determination of optical variations among sensors*
- 3) Test how consistent data various sensors provide across temperature and fDOM gradients, guiding developments towards harmonised sensor networks.*

What types and formats of data will the project generate/collect?

*Mesocosm experiment:
csv files with timestamps*

*Ferrybox integration:
csv files with timestamps and coordinates*

The frequency of the mesocosm experiment is 1Hz and for the ferrybox installation 0.05 Hz.

How is the original data gathered on the TA installation(s) and how do you plan to transfer

it to you?

*In mesocosm experiment, using (python-based) datalogging systems of facility owner
In addition, during calibration, some sensors may collect data using their own software
Packages.*

*In ferrybox, either using the datalogging systems onboard, or extending the system created
for mesocosm experiments.*

What processing on the raw data do you plan? Please differentiate between data quality assurance (handling of outliers, missing and suspect values, null observations) and data harmonisation (code, label and QC flag explanations, consistent use of headers and units, data formatting, conforming to standards).

*Data QC includes outlier detection (based on normal distribution of data in experiments),
experimental failure detection (based on log-book notes during experiment), detection of
suspect values (based on sensor comparison, also in ferrybox)*

*For data harmonisation, we aim to have all sensors logging at same intervals, with same
averaging setups. As data should be in similar csv files, the headers and filenames need to
detail the sensor.*

When do you plan to perform these processing steps by means of a concise planning?

*Data QC is done and closely followed during the experiments, as we need to halt operations
in the case of sensor failures. The same applies in ferrybox deployment, we need to follow
data quality at least weekly, allowing us to react once the ferry is in harbour (every four
days), to decide on maintenance actions or sensor swapping.*

When do you plan to perform these processing steps?

See above.

What is the expected size (Megabyte to Terabyte range) of the data? Please note that in case of multibeam or imagery data, the derived products (such as grids) should be submitted instead of the raw measurements.

We are expecting up to 1 GiB of raw data per month with all sensors present.

To whom might your data be useful, outside your project?

For the scientific community and government monitoring agencies.

2. Making data FAIR

2.1 Making data findable, including provisions for metadata

The **AQUARIUS Data Management approach** is aimed at inclusion of your collected/generated data in leading European aquatic data management infrastructures such as **SeaDataNet** (physics, chemistry, geophysics, geology, and biology), **EurOBIS** (marine biodiversity), **ELIXIR-ENA** (biogenomics), **ICOS-Ocean** (carbon), and **Copernicus INSTAC** (Near- Real-Time data). Inclusion will be mediated through expert data centres which are partners in the AQUARIUS project. They will give training and education to the TA teams about planned data management practices as part of AQUARIUS

WP5. In addition, as part of AQUARIUS WP6, they will give support to the TA teams for the adoption of the preferred metadata and data standards, and ingestion of the resulting metadata and data in repositories that are nodes of the indicated European data management infrastructures. This way, the metadata and data from the TA projects will become FAIR and also available for EMODnet, Copernicus Marine, DTO developments, EOSC, and global data exchanges such as GEOSS and the digital ocean ecosystem which is being developed in the framework of the UN-IOC-Ocean Decade programme. In the AQUARIUS Data Management approach, it is recommended to make use of the online **Data Submission service of EMODnet Data Ingestion and/or SeaDataNet SeaNoe service** for transfer of your metadata and data packages to the AQUARIUS expert data centres.

Paragraph 5.4 of the **AQUARIUS Data Management approach** document lists the minimum set of metadata that EMODnet Ingestion requests you to enter in order to document your data packages submission. In addition, Annex 1 of the document gives a list of items that should be documented during the TA data acquisition and processing activities as these are later needed, when transforming data to the standard metadata and data formats as required.

What naming conventions for your data files will you follow?

The project will follow the naming conventions of the data repository chosen.

Will you in principle be able to follow the **AQUARIUS Data Management approach** for metadata? What possible challenges do you foresee?

Yes, this is fine. Need to note that the experimental and calibration data is not necessarily to be included in abovementioned data repositories. For Ferrybox fDOM data, we would very much like to discuss how to provide consistent metadata (and data) from our project, using several sensors and when applying several correction factors (to make data coherent) and how such novelties in sensor metadata should be treated. The selected data repository preferred would be SeaDataNet, since it is a delivery of reprocessed delayed mode data. The vocabulary used would follow the NERC Vocabulary Server: (<http://vocab.nerc.ac.uk/collection/P01/current/QSCDOMX1/>)

How will you document all this information before submission, especially lineage information (i.e. processing and QC steps)?

This is very essence of our project. We provide sensor (and site) specific correction functions (defined by the experimental data), and the raw data collected by sensors in situ need to be transformed using these correction functions. We need to secure that the metadata information of the use of correction functions always follows the data itself. Actual method needs to be defined (group includes several ferrybox data providers and a Baltic node for Copernicus ferrybox data – so we have quite good standing point to advance this topic with AQUARIUS data-team).

2.2. Making data openly accessible

The AQUARIUS project follows an **open data access policy**. Therefore, the **AQUARIUS Data Management approach** is aimed at inclusion and publishing of your

collected/generated data as open metadata and data in trusted repositories, within the lifetime of the AQUARIUS project. However, temporary embargos (max 2 years after collection) are possible for specific data that require additional processing and to allow for scientific analyses that will lead to scientific papers.

Will you collect/generate data that might require a temporary embargo? Indicate which type of data and motivate its temporary embargo for each case.

The experimental data, including ferrybox data on fDOM, is collected to meet the needs of sensor tests and should not be considered as environmental data. Analysing the data requires deep knowledge on the experimental conditions. We therefore consider it is wise to publish such data only with appropriate explanation (peer reviewed paper) and consider publishing it along the publication process, likely 1 year after the experiment.

Do you also plan to make the data and metadata available on another repository next to the AQUARIUS repositories, for instance an institutional, national or general data repository?

Depending on the publication, we aim for any generally used data repository (Zenodo or Seano)

In case of use of exotic data formats for original data, what methods or software tools are needed to read and understand the data and could you provide documentation about that (including possible open-source code)?

NA.

2.3. Making data interoperable

The **AQUARIUS Data Management approach** is aimed at making the collected/generated data interoperable by adopting common standards from leading European data management infrastructures. Use will be made of standard formats, QA-QC procedures, and controlled vocabularies for many metadata tags. For this process the TA teams will be given support by the AQUARIUS expert data centres, who also can help with use of common tools and services. To be able to do these transformations from original metadata and data to standardised metadata and data, it is important that relevant information about the data collection/generation is noted down during the actual events.

This consists of the **AQUARIUS TA Data Summary Log** which should be maintained by the PI of the TA project scientific team to keep an overview and index of the data collection events. It should be shared afterwards with the AQUARIUS data centre experts as it will provide an index to the collected/generated data and relevant documentation, such as who, when, where, how, etc. It contains only metadata and no data. For keeping this log, the PIs of the TA projects will receive a tablet from WP6 with a preloaded app, that will facilitate entering, browsing, and exporting information. The tablet should be returned to the WP6 team after finalisation of the TA acquisition activities. The **AQUARIUS TA Data Summary Log** will serve as a list for the expert data centres to know what data to expect from where and who and as a checklist for the following steps in the data management workflow.

In addition to the TA Data Summary Log it is recommended that the researchers of the TA teams document several aspects of the data acquisition and following processing, so that the later transformation to the common standards of the European marine data management infrastructures will be easier. Annex 1 of the AQUARIUS Data Management

approach document gives a checklist which documentation TA researchers are recommended to maintain for this purpose.

Will you in principle be able to organise your TA team during the TA activities in order to gather the input required for the **AQUARIUS TA Data Summary Log**? What possible challenges do you foresee?

With experimental data we use our own logbook (including mostly verbal notifications, screenshots, photos etc.) to include all events which may not be useful to include in AQUARIUS TA Data Summary Log, so we do not see this feasible. When we continue with continuous records of ferrybox data, in principle yes, but we need to better understand how this may improve the operations already done by facility owner.

How are you going to organise gathering the input required for the AQUARIUS TA Data Summary Log during the TA activities?

The meta data and sensor outputs are organised by the mesocosm host SYKE and delivered to the PI in order to add selected parts to the AQUARIUS TA Data Summary Log.

2.4. Increase data re-use (through clarifying licences)

The AQUARIUS project follows an **open data access policy**, and the AQUARIUS Data Management approach is aimed at inclusion and publishing of your collected/generated data as open metadata and data in trusted repositories, within the lifetime of the AQUARIUS project. The preferred licence for use of AQUARIUS data is CC-BY-4.0.

Do you foresee any difficulty for your collected/generated data to follow this policy?

No!

If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

NA

3. Other research outputs

Do you plan to generate other research outputs next to the collected/generated data such as digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.) results?

Facility owner needs to create some software to integrate sensors in the AquaBox software and to do primary QC of data. In principle, they are openly available and done using open software.

4. Allocation of resources

Terms & Conditions of AQUARIUS Funding require TA teams to follow the AQUARIUS Data Management Approach. This includes efforts for:

- updating their initial **Data Management Plan (DMP)** per TA project, as included in their original proposal submission, after being awarded and getting ready for the actual TA project deployment
- taking part in WP5 data management training activities to learn about AQUARIUS data management standards and getting experience with tools and services

- using the app during TA activities to complete the **AQUARIUS TA Data Summary Log**
- completing a Cruise **Summary Report (CSR)** in case the TA project will involve scientific cruises with Research Vessels
- transforming the original metadata and data to the AQUARIUS prescribed standards, performing QA-QC, and making these FAIR and ready for inclusion in AQUARIUS repositories, thereby supported by AQUARIUS expert data centres.

Who will be responsible for data management in your TA project?

Sebastian Ehrhart, Data engineer at SYKE.

Anna Willstrand Wranne, project PI.

5. Data security

The operators of the repositories as used for the AQUARIUS archival and long-term stewardship have security provisions in use such as for backups, secure storage, firewalls, etc.

What provisions are in place by the TA teams for data security (including backups, secure storage and transfer) in the initial phase, before the data is transferred to the AQUARIUS expert data centres?

Experimental data is stored either at facility owner Syke server or using Finnish national computing service CSC services (from where it can be downloaded by user group), both with secured backups. Ferrybox data is stored in facility owners' systems (Syke, Niva) with backups, and copies are transferred to users when needed (e.g. using CSC B2Drop service).

6. Ethical aspects

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing?

No

7. Other issues

Will you follow any other national/ sectoral/ departmental procedures for data management, on top of the infrastructure and procedures that AQUARIUS will provide as part of its data management approach?

Facility operator Syke is following Data Management Policy of national marine RI (FINMARI). User group partners are mostly from universities or from governmental institutes, and follow they own policies as well. All of them are in line with FAIR data policy.