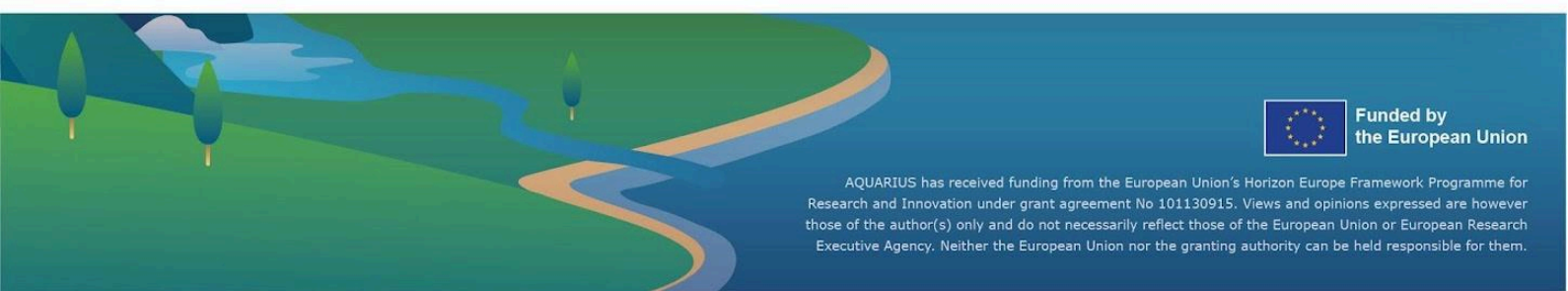




AQUARIUS initial Data Management Plan template for TA Call applications

September 2025



AQUARIUS initial Data Management Plan template for TA Call applications

As part of the Transnational Access (TA) projects in AQUARIUS, many new data sets in a large variety of data types will be collected by the TA scientific teams, making use and combining multiple and different observation installations as provided. There will be a strong effort in AQUARIUS to get the maximum return of investment from the TA activities towards serving the EU Mission and AQUARIUS Partnership targets and associated initiatives and projects with new data, data products, and scientific knowledge. Therefore, AQUARIUS has adopted an **open data policy**, which will be implemented with a **dedicated Data Management approach**, to ensure that all gathered and generated metadata and data will be managed in line with the FAIR principles (Findable, Accessible, Interoperable, Reusable). The metadata and data should become part of the archives managed and operated by leading European data management infrastructures for quality assurance, long term stewardship, and wide access and use.

TA scientific teams are advised to take note of the AQUARIUS Data Management approach document and guideline which can be found at:

<https://zenodo.org/records/13642869>

The AQUARIUS data management flow scheme includes a number of steps from planning to training to deployment to publishing, and a number of instruments, which should be applied in those steps by the TA scientific teams.

An initial step is that TA Call applicants are requested to complete an **initial Data Management Plan** for their proposed project, following the **AQUARIUS DMP template phase 1**, and to submit this as part of their TA Call application

Once the TA proposal has been awarded, the TA project team will be requested to review and refine the DMP once the TA team has a better understanding of the TA project that will be deployed and the AQUARIUS data management approach. AQUARIUS expert data centres can assist with the Call process and respond to questions about the AQUARIUS data management approach, for as far as answers cannot be found in the AQUARIUS Data Management approach document. In the phase after the TA project awarding there will be a kick-off meeting for selected TA projects with the AQUARIUS expert data centres to give support for completing and refining the TA project DMP template.

The questions in the DMP template are derived from the Horizon Europe DMP template and adapted for use as part of the AQUARIUS data management approach.

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

1. Data Summary

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

The purpose is to collect, integrate and analyse ecological, hydrological and biogeochemical data to explain the spatial distribution and post-settlement survival of calcifying benthic species (principally *Leptogorgia sarmentosa* gorgonia) on *El Biòtop* artificial reef. Specifically, nutrient concentration data (nitrate, phosphate, silicate) will be used to identify nutrient hotspots and environmental conditions that favour establishment and persistence of these species. This directly supports the scientific objectives: (1) characterise species spatial distribution by depth and substrate; (2) determine environmental drivers (including nutrients) of that distribution; (3) communicate about evidence-based artificial-reef design.

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

Several types of data are concerned in the project:

1. Ecological data: species presence/absence, abundance estimates, percent cover, GPS coordinates, depth, substrate type.
2. Nutrient lab data: nitrate, phosphate, silicate concentrations from autoanalyzer outputs.
3. Metadata and documentation: sampling protocols, instrument calibration files, README files, QC logs, and codebooks (text/PDF). Analysis artifacts: R/Python scripts, statistical outputs, model parameter files, and GIS project files

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

Field team will collect *in-situ* water samples at *El Biòtop*. Nutrient samples will be chilled and transported following cold-chain protocols to OOB (primary analytical site) for analysis by train or by car. Analytical data produced at OOB will be exported as CSV/NetCDF.

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).

All data will follow a strict QA/QC protocol to ensure accuracy and comparability. Instruments will be calibrated against certified standards before and after campaigns. For nutrient analyses, blanks, standards, and duplicates will be used, with detection limits reported. Outliers ($>3\sigma$) will be flagged and verified against field logs, while missing or null data will be coded as NA (not measured), ND (not detected), or sensor gaps, with causes documented. Photographic data will be checked for GPS metadata, timestamps, and image quality, removing corrupted files.

Data harmonisation will include standardised units (μM for nutrients) and consistent headers across all datasets. Controlled vocabularies (Darwin Core, ISO) and standard QC flags (SeaDataNet/WOCE) will be applied. Time series will be converted to UTC and harmonised to consistent temporal resolutions (monthly averages) with transparent interpolation. Final nutrient dataset will be stored in NetCDF format following CF conventions to ensure compatibility with AQUARIUS and international marine data repositories.

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

The processing steps are planned to be done between March and April 2026 as part of her Master's internship in Marine Science. This date and OOB location is convenient because can be easy reachable by car or train from the artificial reef. Data will be processed during the next two months in order to include it in the Master Thesis and start writing a scientific publication.

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**¹ (physics, chemistry, geophysics, geology, and biology), **EuroBIS**² (marine biodiversity), **ELIXIR-ENA**³ (biogenomics), **ICOS-Ocean**⁴ (carbon), and **Copernicus INSTAC**⁵ (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**⁶ and/or **SeaDataNet SeaNoe service**⁷ 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.

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

Yes. We will adopt AQUARIUS metadata and submission guidelines (use of AQUARIUS TA Data Summary Log and metadata templates) and follow recommended standards (INSPIRE, ISO19115, DarwinCore, NetCDF-CF). The expected size of the data is one megabyte. This dataset might be useful for other marine researchers interested in a one-round year series of nutrient concentrations next to gorgonian colonies in Torredembarra, a site in the NorthWest Mediterranean influenced by the NorthCurrent (NC) originated in the Gulf of Lions (France).

¹<https://www.seadatanet.org>

²<https://www.eurobis.org>

³<https://www.ebi.ac.uk/ena/>

⁴<https://www.icos-cp.eu/observations/ocean/otc>

⁵<https://marine.copernicus.eu/about/producers/insitu-tac>

⁶<https://www.emodnet-ingestion.eu>

⁷<https://www.seadatanet.org/Software/SEANOE>

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

Every dataset will be accompanied by a comprehensive metadata file and a processing log describing: raw file name → processing step → script used (with version) → processed file name. QA logs (standards, blanks, instrument calibration sheets) will be archived with datasets.

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.

A maximal-2-year embargo after finalised QA is planned to allow manuscript submission and institutional review. In a one-year round series data processing can take one year, as usually all samples are analyzed at the same time and the first samples are frozen in the meantime. In addition, environmental data needs additional processing, possibly taking six months since it is collected. Ecological data can be processed within weeks, but will be out of context without the nutrients and environmental data. As such, we estimate that a maximal 2-year embargo will be needed to compile all datasets, subscribe a manuscript and create a clean dataset for public use. This additional processing can be conducted by Paula Isabel Hueso Jiménez as Early Career Research and shared with the project team, especially other early career researchers that could use it for their PhD.

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?

The collected data will be deposited in the AQUARIUS repositories under a restricted license until it is clean to go public and the related PhD is finished. Once finalized, the data center will change the licence to an unrestricted one (public access). In addition, data and metadata will be deposited in OOB and Universitat de Barcelona institutional repositories.

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?

Yes. TA team will organise designated roles (field lead by Irene Gonzalo Cruz, data analysis by Paula Isabel Hueso Jimenez, OOB contact to help analyse nutrient data with Paula Isabel Hueso Jimenez . We will complete the AQUARIUS TA Data Summary Log when collecting *in situ* nutrient data and describe post-processing.

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 major difficulties

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?

The collected nutrient data could be used to optimize protocols regarding environmental monitoring in the area.

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)**⁸ 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.

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?

All data transfers to project or institutional servers will be carried out using secure protocols (SFTP over SSH) or certified institutional transfer services. Access to all storage systems will be restricted to authorized personnel through institutional user accounts, implementing two-factor authentication whenever available.

A verified copy of the final, quality-controlled datasets and metadata will be archived on the OBB institutional long-term storage system before submission to the AQUARIUS or other designated data repositories, ensuring both data integrity and long-term preservation.

6. Ethical aspects

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

Outreach activities will be designed to be inclusive and to provide feedback and benefits to local communities and stakeholders. All these issues will be explicitly addressed in metadata and data use statements.

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?

The GenConMed project will comply with national and institutional research data management policies in both Spain and France, ensuring that all data are archived, curated, and shared in accordance with the principles of open access and long-term preservation. Data handling and archiving will follow the open data mandates established by Spanish and French research agencies, which encourage deposit in certified repositories and adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles.

⁸<https://www.seadatanet.org/Metadata/CSR-Cruises>

