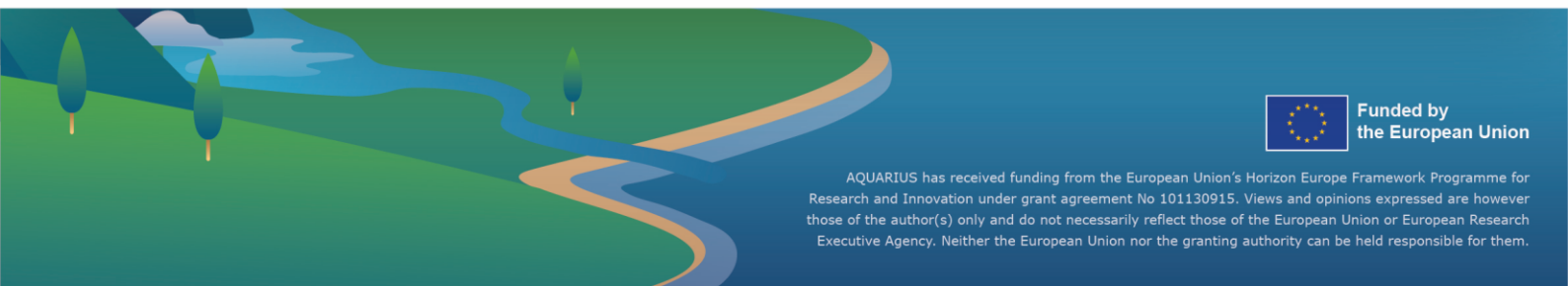




AQUARIUS initial Data Management Plan template for TA Call applications

September 2024



AQUARIUS has received funding from the European Union's Horizon Europe Framework Programme for Research and Innovation under grant agreement No 101130915. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

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

The purpose of the data collection/generation in the **Hydro-MED** project will be to better understand the integrated hydro-morpho-eco dynamic processes within transitional waters, such as estuaries and deltas, and the key roles they play in the accumulation, transfer and dispersal of anthropogenic-derived marine pollution, by acting both as a major pollution source and sink in the coastal marine environment. These data will be collected via targeted field campaigns at the Ebro Delta Supersite, conducted over an extended period (15 June 2026 – 14 May 2027), with additional data generated for the site from complementary numerical modelling simulations.

The field data will consist of individual raw data files (e.g. *.dat; *.txt) obtained from each of the scientific instruments utilised during the field campaigns, including: acoustic doppler current profilers (ADCPs/ADVs); optical backscatter sensors (OBS) for suspended sediment measurements; conductivity-temperature-depth (CTD) probes for salinity/density profiling; multibeam echo-sounder for sediment bed morphology measurements; conventional meteorological sensors (i.e. wind speed/direction, rainfall); and pressure meters (i.e. sea level gauges). Lagrangian buoy drifters will employ GPS sensors to track their position from release (i.e. source-to-sink). Data from the rhodamine dye tracing experiments will combine direct in-situ concentration measurements (via fluorimeters) and high-resolution timelapse aerial images captured from UAVs (i.e. drones). Other field data collected at the supersite will include water and bed sediment samples to analyse pollution levels and identify the presence and composition of microplastics and other marine litter [e.g. via lab analysis using Fourier-transform infrared spectroscopy (FT-IR) and micro X-ray tomography (micro-XRT)].

Outputs from the numerical model (ROMS, BOM) simulations will include datasets (e.g. NetCDF files) on the prevalent hydrodynamics within the wider delta region; exchange flow & salt-wedge dynamics in the estuary/river mouth; coastal sediment transport processes & evolving morphodynamics; and the transport, dispersion & accumulation of pollutants (including microplastics and marine litter) in the dynamic coastal system.

The original data will be gathered via instrumentation (see above) deployed at the Ebro Delta supersite. Instrumentation that is deployed over extended periods (e.g. ADCP; buoy drifters) will be monitored on a daily basis via telemetry to ensure functionality and data acquisition are occurring normally. Data from these deployments will then be downloaded once the instrumentation is collected/retrieved from the site. Other datasets from portable instruments (e.g. ADV, OBS, CTD, echo-sounder, fluorimeters, etc.) and equipment (e.g. UAVs) utilised during the field campaigns will be downloaded daily.

Much of the raw data collected from the field campaigns will be processed and analysed using standard data processing software such as MATLAB, Python, Excel, or R. Post-processing of ADCP data (for example, to determine velocity profiles and turbulence characteristics induced by currents and waves) will also utilise device-specific data processing software. This processing will include checks on data quality (e.g. correlation coefficients) and the use of appropriate filtering techniques to detect and correct data outliers. The raw aerial image data obtained from UAVs will undergo advanced image post-processing using computer vision techniques (e.g. intensity thresholding, edge detection) to enhance spatial and temporal resolution, enabling detailed analysis of tracer dispersion plumes and regions of accumulation to be detected and assessed. All processed datasets will be harmonised to ensure consistency in the use of standard headers and units, QC flag explanations, data formatting, and presentation, with generation of appropriately detailed metadata (including direct links to the raw datasets) also utilised to aid data discovery and reuse (see 2.1). As the raw data will be collected over an extended period (15 June 2026 –

14 May 2027), during several field campaigns, the data will be processed and analysed in the immediate aftermath of each campaign. This is required to inform future campaigns (see below) and will allow data to be sent to data centres for publishing in AQUARIUS repositories in a timely manner (before the end of the project).

Defined protocols for data management, storage, harmonisation, and processing will be agreed upon by the user group and providers from the outset of the project to ensure it is undertaken in a consistent manner and conforms with appropriate standards. All initial data processing will be undertaken as soon as the individual datasets have been collected/retrieved from the site, both to ensure data quality & consistency and to enable the informed planning of the future field campaign periods.

The processed data will have an expected overall size in the range 500 Mb – 1 Tb. It is anticipated that the processed data will be useful to coastal scientists and engineering practitioners working in river-estuarine-marine hydrodynamics, coastal sediment transport dynamics and morphodynamics, pollution and water quality, and the protection of sensitive coastal ecosystems and habitats.

2. Making data FAIR

2.1 Making data findable, including provisions for metadata

The user group, working closely with the Supersite provider, intend in principle to follow the **AQUARIUS Data Management approach** for metadata to facilitate the inclusion of the collected/generated data in the **Hydro-MED** project within leading European aquatic data management infrastructures (e.g. SeaDataNet; Copernicus). This will require preparation of a standard list of metadata elements when datasets are being submitted via the online Submission service of EMODnet Data Ingestion. In addition, the user group will follow the checklist for required information (given in Annex 1 of the AQUARIUS Data Management Plan) to document data acquisition and processing during the **Hydro-MED** project. This includes required information on the TA installation platform and instrumentation used for data collection and sampling; measurement locations and coordinate referencing system; time, depths, and locations at beginning (and end) of sampling/measurement windows, etc. All this information will be initially recorded in logbooks but will then be transferred into electronic metadata files (e.g. *.dat; *.txt) for secure storage and data accessibility. It is proposed that all data files will follow strict naming conventions as follows: <ProjectName>-<Measurement Type>-<GeoRef>-<Date> (e.g. HydroMED-ADCP-Lat_Long-DDMMYY). Metadata for the data sets collected at the site will include georeferenced maps showing the precise locations of measurements.

2.2. Making data openly accessible

All data that will be collected/generated during the **Hydro-MED** project will follow the FAIR principles of Findability, Accessibility, Interoperability, and Reuse of digital datasets. Temporary embargos, specifically associated with release of the raw and processed data collected during the field campaigns planned in **Hydro-MED**, may be requested (for up to 2 years after data collection) to facilitate further data processing and analysis that will lead to scientific publications, prior to the data being made available to the public. Where this embargo is likely to extend beyond the end of the AQUARIUS project, data will be deposited in AQUARIUS repositories under a restricted license (no public access) until the publications have been finalised. After this point, the data centre will change the license for the data to an unrestricted one (public access). In cases where raw data sets include image formats (e.g. *.tif, *.jpg files), Matlab or Python codes used to process and analyse the raw images will be provided (as open-source code), with accompanying documentation outlining these image processing and analysis procedures. Raw and processed data sets, along with associated metadata, will also be stored in data repositories at UPC (infrastructure provider) and the University of Dundee.

2.3. Making data interoperable

In order to ensure that all data that is collected/generated during the **Hydro-MED** project is interoperable, the PI and TA project user group, with the support of the Supersite provider, will ensure that all relevant information is noted down during the project activities at the Supersite through the initiation and continual update of the electronic **AQUARIUS TA Data Summary Log** in order to maintain an up-to-date overview and index of the data collection events undertaken at the Supersite. At each of the individual field campaigns planned during the **Hydro-MED** project, a delegated member of the TA project user group participating in the field work will be responsible for updating the log via the preloaded app. In order to ensure that this information remains consistent throughout the planned field campaigns, all members of the TA project user group will be shown how to collate the required metadata for the data collection events and enter it via the preloaded app. Between individual field survey periods, the tablet containing the summary log will remain in the possession of the Supersite provider, with information also stored securely on the UPC data repository.

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

The user group do not foresee any issues with the AQUARIUS open data access policy for the inclusion and publication of data collected/generated during the **Hydro-MED** project as open metadata and data within trusted repositories, within the lifetime of the AQUARIUS project (subject to any user requested embargoes of up to two years from data collection). Any requested embargo will not prevent the inclusion of Hydro-MED data in the AQUARIUS repositories during the embargo period under a restricted license, reverting to an unrestricted license at the end of the embargo period.

3. Other research outputs

As a planned output from the Hydro-MED project, on top of expected journal papers and conference proceedings, we plan to utilise field monitoring datasets and complementary multiscale numerical modelling to develop a simplified engineering tool, based on internal hydraulic modelling theory. This will provide enhanced physical understanding on river-sea connectivity, and the specific conditions for water (and pollutant) retention, exchange, and flushing times within the delta region, providing a unique delta water quality management and pollution monitoring tool. It is anticipated that this tool will also be applicable to model water and pollution exchange in a wider range of estuarine/deltaic case studies under different hydro-morphodynamic conditions.

4. Allocation of resources

The TA user group accepts the terms & Conditions of AQUARIUS Funding requiring 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)**¹ in case the TA project will involve scientific cruises with Research Vessels (not applicable for the **Hydro-MED** project),
- 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.

Overall responsibility for data management in the Hydro-MED project will be shared jointly between the project user group (led by Dr Alan Cuthbertson, Hydro-MED PI) and the research infrastructure provider (led by Dr Manel Grifoll, iCIEM Ebro Delta Supersite).

5. Data security

All data (i.e. collected, generated and/or post-processed) and associated metadata will also be stored securely within a data repository at UPC (the Supersite provider). Remote access for the TA project user group to this data, as well as the corresponding numerical model simulation results, will be provided via OPeNDAP and the secure file transfer capacity of the Supersite provider. Any post-processed data generated at the external TA user group institutes will also be uploaded to UPC data repository (along with accompanying metadata), with back-ups held within their own secure institutional data repositories, prior to all data being transferred to the AQUARIUS expert data centres.

6. Ethical aspects

To the best knowledge of the PI and TA project team, there are no ethical or legal issues that would impact on data sharing amongst the participants and participating institutes.

7. Other issues

The PI and TA project team agree to follow any other national/sectoral/departmental procedures for data management, on top of the infrastructure and procedures that AQUARIUS provide as part of its data management approach.

¹<https://www.seadatanet.org/Metadata/CSR-Cruises>