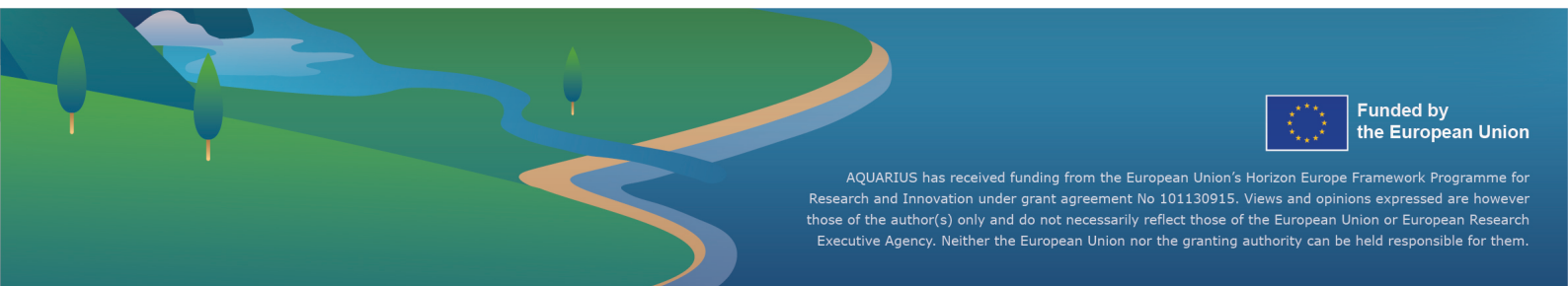




AQUARIUS Data Management CETAIR

October 2026



Funded by
the European Union

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

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

The project operationalises UAV-based line-transect distance sampling for small delphinids in northern Portugal (Aguçadoura/Viana do Castelo) and validates AI-assisted detection/tracking against concurrent vessel surveys. Data collected will (i) derive unbiased density and abundance estimates from UAV imagery; (ii) quantify detectability drivers (altitude, GSD, viewing angle, sea state, glare); and (iii) benchmark UAV- vs vessel-based estimates, culminating in a Best-Practices & Operational Guidance for cetacean monitoring in offshore renewables contexts.

Data generated by the project will contribute beyond its main objectives. UAV imagery, telemetry and flight performance data may support robotics and remote sensing researchers in assessing and optimising UAV survey performance under marine conditions. Annotated imagery, detection records and tracking outputs may be reused by AI researchers to develop, train and validate object detection and tracking methods, which can be extrapolated to wider marine species than small delphinids. The harmonised ecological, environmental and survey effort datasets will also support marine ecologists, conservation researchers, environmental consultants and public authorities in improving evidence based marine management, namely for assessing species distribution and developing monitoring and mitigation approaches for areas requiring stricter protection.

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

UAV flights will generate raw MP4 video files and JPG still images, along with flight logs (.txt, .DAT, .csv). Processed data will include effort and transect geometry, environmental covariates, perpendicular distances, and encounter counts for distance sampling analysis, which will be used to compare with boat based visual detections. The expected size of the processed data is between 5-100 GB.

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

UAV platforms collect video and imagery over predefined transects, with synchronised logs containing telemetry and camera data, which are offloaded on site to SSDs and backed up to WavEC's cloud storage and NAS service once in the office.

- 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).

Raw data will be thoroughly verified for quality assurance and quality control, namely time syncing imagery, completeness checks, recording on sanity checks, visual quality control for blur.

The most important parameter derived from raw data is distance, or any secondary parameter that allows to calculate distance (e.g. GSD, altitude, camera angle, etc). Thus, any outliers or missing values in this parameter will be excluded from analysis following the distance sampling methodology. Nonetheless, these videos will still be utilized for calibration checks and for assessing AI model performance/model development. Missing observations and outliers for the remaining parameters can be handled statistically.

Harmonisation of data will be achieved by preparing templates and checklists for data processing, quality assurance and control, and analysis. These will ensure to unify time to UTC as per ISO-8601, spatial and geolocation to UTM-WGS84 (EPSG:4326), use standard column names and units, use a defined vocabulary, and consistent species coding.

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

Following each survey (<1 week), collected data will undergo quality assurance and control, flight logs, telemetry and camera data will be extracted. Between each survey (1 week – 3 months), data will undergo preliminary collation and processing ensuring harmonisation, including effort and transect geometry, environmental covariates, perpendicular distances, and encounter counts. Interim distance analysis for abundance and density estimates will be conducted to highlight any hidden issues associated with data collection methodology. By Q1 2027, final harmonisation and distance analysis will be conducted incorporating results from both surveys.

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.

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>

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. The team will follow the AQUARIUS metadata approach. The main potential challenge is contractual coordination with the private developer (CorPower Ocean), but this is not expected to block compliance given the EU-SCORES context and its emphasis on dissemination. Where needed, we will keep metadata openly available even if files are under a short embargo, and apply redaction/coarsening for sensitive locations

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

The team will make a small "Submission Pack" and keep it with each dataset. It will include a metadata file (what the data are, where/when/how they were collected, who did it, which files are included, and how to cite), a step-by-step methodological note, and any key settings we used. Additionally, it will include quality checks summary for missing files, timestamps, etc., and software versions used.

Before submission, a quick completeness checklist will be conducted to make sure this information is present. If we blur/round sensitive locations or remove anything for privacy, we'll say so clearly in the metadata file and keep the original files safely archived.

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.

Derived, non-sensitive data (e.g., transect/effort geometry, sightings detection summaries and flight parameters) will be transferred to an EU infrastructure within 2 months of data collection.

The final dataset, including post-processed videos and images, annotations and model outputs, will be reviewed prior to deposition to identify and appropriately manage any sensitive or confidential information. Where necessary, such information (e.g., identifiable vessels, persons or confidential project locations) will be removed, anonymised or generalised before the dataset is deposited in SEANOE. The curated dataset will then be made available under a temporary embargo of up to 2 years to allow completion of the primary analyses and preparation of a peer-reviewed publication. Following the embargo period, the dataset will be made openly accessible in accordance with the AQUARIUS open data policy.

- 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?

Yes, data will also be made available at the Portuguese National Oceanographic Data Centre (NODC <https://nodc-portugal.pt/#about>).

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?

A delegated member of the TA user group will be responsible for ensuring that all data collection events will be completed on the Aquarius TA Data Summary with the required inputs and no challenges are foreseen. The TA team is a small team, from only one organisation that will coordinate activities with the hosting organisation, facilitating communication and logistics to get information across. Additionally, much information required to be documented during data acquisition is automatically recorded in the flight logs.

Foreseen challenges and mitigations include:

- Limited connectivity offshore: logs will be acquired offline and syncing when ashore.
- Time pressure during manoeuvres/battery swaps: the team will have a dedicated scribe not involved in flying, and allow voice notes for later entry.

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 are expected. Our datasets are primarily original and can be released under CC-BY-4.0.

Possible challenges may arise with coordination with CorPower Ocean. However, we'll ensure the collaboration agreement explicitly allows open metadata and CC-BY release of our data after embargo (2 years). The survey area will also cover a larger area than that where CorPower's device is installed. Additionally, sensitive content (precise locations, identifiable vessels) will be redacted/rounded before release and document changes in the metadata file.

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?

Yes, the team plans to use the acquired images to create training and validation datasets that will allow to further develop its current AI model for dolphin detection.

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

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

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?

Data management of the TA project will be conducted by the PI organisation, WavEC. WavEC implements 3-2-1 backups for data storage, with primary storage on encrypted NAS, secondary storage encrypted external HDD, and third storage off-site mirror from NAS.

6. Ethical aspects

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

No.

Animal welfare is considered during flights, maintaining altitudes and profiles designed to minimise disturbance, and operations will be halted if disturbance is observed. All data will be acquired with the necessary permits, including BVLOS/SORA authorisations. Additionally, operations will be conducted outside exclusion/MPA zones.

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?

Yes. In addition to AQUARIUS, we will apply:

- EU/Portugal data protection rules (GDPR). We don't expect personal data, but if any is incidentally captured, we will anonymise/blur and document it.
- Operational record-keeping: retain flight and safety logs required for aviation/maritime authorities (ANAC/AAN/AMN/port authority) alongside our data package
- Contract with CorPower Ocean: ensure agreements allow open metadata from the start and CC-BY release of our data after any short embargo