

AQUARIUS initial Data Management Plan

ABACUS 2026

This data management plan ensures alignment with the objectives outlined in the AQUARIUS Deliverable 6.1 - Data Gaps Report. By providing high-resolution, multi-parameter ocean observations in the Algerian Basin, this project addresses critical data gaps related to:

- **Marine ecosystem health:** Through the monitoring of physical, chemical, and acoustic indicators, enabling a better understanding of oceanic processes and their impacts on biodiversity.
- **Pollution assessment:** By fostering passive sampling strategies to detect and map emerging pollutants, contributing to efforts for pollution-free seas as per the European Mission Ocean 2030.
- **Eddy activity and physical variability:** By combining glider data with satellite altimetry and SWOT missions, the project enhances knowledge of mesoscale and sub-mesoscale dynamics, crucial for regional ocean modelling.

This data management plan is aimed at curation, preservation and description of collected data flows from SOCIB gliders to the main European marine data portals (such as Copernicus Marine, SeaDataNet and EMODnet). The proposed approach emphasizes alignment with key global stakeholders while adopting Ocean Best Practices (OBPs) throughout the entire data value chain, aiming to enhance the impact of research results in addressing future challenges. In particular, the proposed plan adheres to the recommendations of the AQUARIUS Data Management Approach and of the EuroGOOS Data Management, Exchange, and Quality (DATAMEQ) Working Group, towards greater consistency and interoperability within the marine data community.

To this aim, the proposed Data Management Plan also takes advantage of the previous experiences of the ABACUS 2026 PI and scientific team, gathered in the framework of several national and international initiatives (e.g. JERICO-S3, JERICO-Next, Horizon, Calypso, RITMARE, MONGOOS, EUROGOOS, SOOP, etc.), with regard to data management and the building of an integrated data system.

1. Data Summary

- **Purpose of Data Collection:**
To monitor physical, chemical, and biological parameters in the ocean using a Slocum deep glider provided by SOCIB glider facility, equipped with an additional passive acoustic recorder and a simil-POCIS pollutant passive sampler provided by the ABACUS team, with a focus on assessing the Algerian Basin's environmental status and contributing to the European Mission Ocean 2030 goals.
- **Data Types and Formats:**
Data will be collected using SOCIB Slocum deep gliders, which are equipped with a suite of sensors able to collect physical (temperature and conductivity), biogeochemical (chlorophyll fluorescence, oxygen, CDOM, PAR, backscatter at 700 nm and turbidity)

and acoustic observations. Standard operating procedures will be followed by the SOCIB glider facility team for deployment, maintenance, and retrieval. For passive chemical sampling, we will use a simil-POCIS, a sampler more robust than the commercial POCIS (Polar Organic Chemical Integrative Sampler), developed at the University of Genoa. These passive samplers are effective for monitoring organic chemicals in various aquatic environments and are capable of adsorbing a wide range of contaminants, including pharmaceuticals and personal care products, pesticides, endocrine disruptors, perfluorinated compounds (PFAS), plasticizers, and their degradation products. Simil-POCIS are easy to use, require no maintenance, and do not need an energy source during the exposure period. As suggested, feasibility of simil-POCIS addition onto the Slocum gliders have been already discussed with SOCIB facility operators.

In summary, the collected data will include:

- Physical and biological parameters collected by glider sensors: Temperature, depth, conductivity, salinity, chlorophyll-a concentration, turbidity, oxygen concentration, oxygen saturation, derived ocean currents.
- Acoustic observations: Passive acoustic sensor (Acousonde) data.
- Chemical parameters: concentrations ($\mu\text{g L}^{-1}$, ng L^{-1}) of pollutants retrieved from simil-POCIS passive sampler.

Data will be collected in standardized formats, including:

- NetCDF for glider's physical and biochemical data.
- MT (or other acoustic formats) for Acousonde acoustic data.
- CSV (or similar) for simil-POCIS data.

- **Expected Data Volume:**

Approximately 50 GB per glider mission, depending on the duration and resolution of the observations.

- **Quality control and processing steps:**

At the end of each survey, SOCIB Data Centre will handle collection of all the raw data stored in the internal memory of the glider, initial storage, glider data curation and preliminary semi-automatic quality control on level 1 and level 2 data. Real-time visualization, data publication on the SOCIB repository, and DOI minting will be also carried out by SOCIB Data Centre.

Delayed mode data processing for level 1 and 2 glider observations will include thermal lag correction, filtering and 1m bin vertical averaging. Level 1 will consist of data regularised, corrected and/or derived from raw glider data. Each variable will be stored as sequences of measurements along the glider trajectory, with interpolated position coordinates to match the times of measurement by the sensor, and with unit conversions and filters applied. Each variable will be described and commented on individually. Level 2 data will include the regularly sampled vertical profiles generated from depth binning of level 1 data. The profiles will be obtained by interpolation of level 1 data to produce regular homogeneous and instantaneous profiles from each up- or downcast, using the mean time and position of the corresponding cast for the profile location and time. In this level, each variable will be stored in a matrix.

Level 1 and level 2 data will be firstly transferred to the ABACUS 2026 science team via ftp and/or restricted repository, then made publically available (as detailed in the following section 2) after validation, assurance (handling of outliers, missing and suspect values, null observations) and quality check (filtering, flagging, basic processing) operations, which make data and metadata compliant for distribution to the main European data portals, within 2 months after the cruise run for physical and biological parameters collected by glider sensors and 4 months for acoustic observations and chemical parameters requiring lab work.

The ABACUS 2026 project team will oversee glider data level 1 and level 2 processing, as well as level 3 production, Acousonde and simil-POCIS data curation and management, interaction with AQUARIUS expert data centres, and data versions updates.

In particular, an additional quality inspection on level 2 glider data will be carried out at Parthenope University (Italy) and IMEDEA (Spain) to identify and discard residual bad data and artefacts after the standard quality control provided by SOCIB Data Center. This procedure will include an additional single-point spike control, the interpolation of single missing data along the profiles, the application of a dedicated median filter, a five-point running mean along the depth and, finally, an iterative comparison between adjacent profiles. This final product is indicated as level 3 and will be freely available upon request within 9 months after the end of each survey.

Passive acoustic data quality check and processing will be based at Université du Québec à Rimouski (Canada). Firstly, raw data acquired by passive acoustic monitoring sampler will be converted into standard formats (e.g., WAV), timestamps in the acoustic data will be synchronized with the glider's data, and QC will be applied. This step includes noise removal (with a special care on the detection and removal of the noise generated by the glider's oil pump during surfacing and bottom inflexion), corrupted data deletion and metadata validation (ensuring that location, depth, and timestamp metadata align with glider deployment logs). Then acoustic signal will be processed through spectral analysis to identify frequency bands of interest (e.g., for marine mammals, fish, wind, rain or anthropogenic noise) and apply specific tools to classify detected signals into categories of interest and specific acoustic events. Acoustic data will be made available via open access repository within one year after the end of each survey, in hybrid millidecade format (Martin et al. 2021).

Simil-POCIS passive sampler data will be processed at University of Genoa, Italy as detailed in the research plan and made available via open access repository within one year after the end of the second survey. Once transported back to the laboratory in Italy, passive samplers will be thawed, the damp sorbent transferred into an empty SPE cartridge with the aid of approximately 20 mL of ultra-pure water and eluted using 20 mL of MeOH and 5 mL of DCM:IPA (80:20 v/v). The obtained eluate will be then reduced to dryness on a rotary evaporator, reconstituted in 1 mL of methanol and filtered through a 0.2 µm hydrophilic PTFE filter. Appropriate dilutions were performed before HPLC-MS/MS analysis, in order to detect and possibly quantify the highest possible number of analytes. Procedure is detailed in MacKeown et al. (2024). Data will be then exploited by all ABACUS 2026 scientists. Since most of the analytical procedures on pollutant passive sampler data are non-conservative, the long-term conservation of the collected samples is not expected.

2. Making data FAIR

2.1 Making data findable, including provisions for metadata

Data will be assigned unique and persistent identifiers (DOIs) upon publication and will be accessible through SOCIB's Data Centre, AQUARIUS repositories, and main European data portals and leading European aquatic data management infrastructures, such as SeaDataNet, which comply with the ISO 19115 and INSPIRE standard. As usual, metadata will be always provided with data following EMODNet and SeaDataNet standards and guidelines, in accordance with AQUARIUS Data Management approach document. Data availability will be also promoted through the Aquarius TNA channels (e.g., making use of the Blue-Cloud Virtual Research Environment).

To achieve this goals, ABACUS 2026 scientists will also take advantage of training and education about data management practices provided in the framework of AQUARIUS WP5 by expert data centres which are partners in the AQUARIUS project.

2.2. Making data openly accessible

Access will align with FAIR principles, ensuring that users can retrieve data without restrictions. Glider level 1 and level 2 data will be made publicly available via SOCIB's Data Portal, AQUARIUS repositories and main European data portals, under an open-access policy within 2 months after the cruise run for physical and biological parameters collected by glider sensors and 4 months for acoustic observations and chemical parameters requiring lab work. Level 3 glider data will be made available within 4 months after the end of each survey, allowing sufficient time for preliminary analysis and additional quality control.

For passive acoustic data and simil-POCIS pollutant passive sampler data, which may require more extensive processing and validation, a temporary embargo of up to 12 months after the end of the project may be applied.

However, if these data are included in a scientific publication prior to the end of the embargo period, they will be released alongside the publication.

As most analytical procedures applied to POCIS samples are non-conservative, long-term preservation of these samples is not expected.

2.3. Making data interoperable

Data and metadata will adhere to international standards, such as CF Conventions for NetCDF files. The use of standardized vocabularies for ocean parameters and pollutants will ensure compatibility across systems; to this goal, we adopt the common formats for metadata and data with the existing marine data infrastructures (e.g., EMODNet, Copernicus Marine Service, and SeaDataNet).

Moreover, in addition to standard documentation of the data acquisition and following processing, ABACUS 2026 PI and scientific team will maintain an AQUARIUS TA Data Summary Log to keep an overview and index of the data collection events. To this aim, ABACUS 2026 team will also take advantage of the support given by the AQUARIUS expert

data centres if exceptional transformations from original metadata/data to standardised metadata/data will be needed.

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

All datasets will be accompanied by clear and detailed metadata, including information on data collection methods, quality control procedures, and usage licenses. Data will be licensed under Creative Commons Attribution (CC BY 4.0), encouraging wide reuse while ensuring proper attribution to AQUARIUS TNA initiative, SOCIB glider's facility and ABACUS 2026 project contributors.

3. Other research outputs

We do not plan to generate other research outputs next to the collected/generated data such as digital or physical products.

4. Allocation of resources

The ABACUS team will follow the AQUARIUS Data Management Approach, as outlined below:

- The current initial ABACUS 2026 Data Management Plan will be reviewed and updated following the project's approval, ensuring it is fully prepared for implementation prior to the project's kick-off.
- ABACUS 2026 PI and scientific team will take part in WP5 data management training activities to learn about AQUARIUS data management standards and getting experience with tools and services.
- ABACUS 2026 PI and scientific team will use the app provided by AQUARIUS during TA activities to complete the AQUARIUS TA Data Summary Log.
- The original metadata and data will be transformed 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

● Short-term Storage:

Data will be stored on SOCIB's servers during the glider mission and post-recovery phase. Archival will be managed through SOCIB's data repository and the University Parthenope central archive, ensuring redundancy and data preservation before the data is transferred to the AQUARIUS expert data centres.

- **Backup:**

Regular physical and digital backups will be maintained in geographically separate locations to mitigate data loss risks.

6. Ethical aspects

The project will ensure that data collection and usage adhere to ethical guidelines, including privacy considerations where applicable.

7. Other issues

No other issues are expected.

8. References

Martin, S. B., Gaudet, B. J., Klinck, H., Dugan, P. J., Miksis-Olds, J. L., Mellinger, D. K., Mann, D. A., Boebel, O., Wilson, C. C., Ponirakis, D. W., & Moors-Murphy, H. (2021). Hybrid millidecade spectra: A practical format for exchange of long-term ambient sound data. *JASA Express Letters*, 1(1), 011203. <https://doi.org/10.1121/10.0003324>

MacKeown, H., Scapuzzi, C., Baglietto, M., Benedetti, B., Di Carro, M., Magi, E. (2024). Wastewater and seawater monitoring in Antarctica: Passive sampling as a powerful strategy to evaluate emerging pollution, *Science of The Total Environment*, 925, 171755. <https://doi.org/10.1016/j.scitotenv.2024.171755>