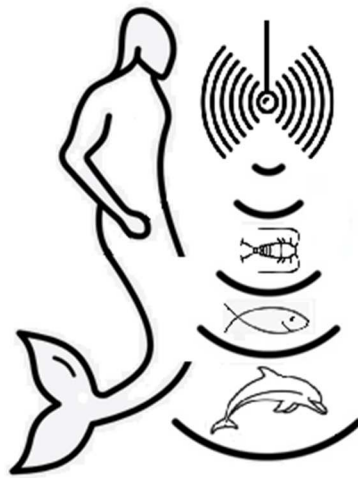
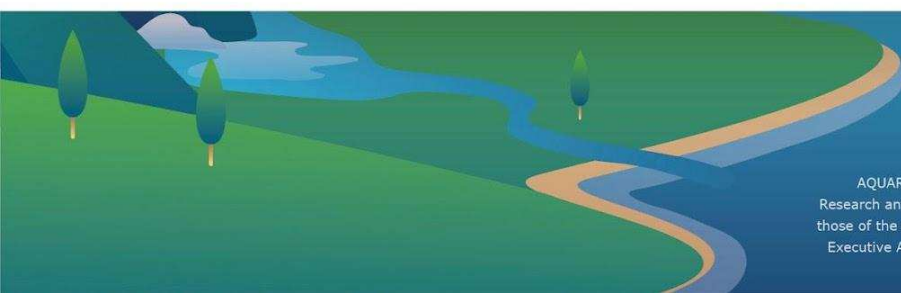




Acoustic Quality and impact of UnderwAter NoIse on Marine Fauna [AQUANIMF]

Initial Data Management Plan



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.

1. Data Summary

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

All data that will be collected during the project, will be processed to improve the knowledge of the acoustic soundscape which is a combination of natural, biological, and human-generated sounds. These acoustic components, known respectively as geophony, biophony, and anthrophony, create a complex and dynamic sonic environment that varies across space, time, and oceanographic conditions and whose impact on fauna is not yet fully known. The acoustic data both from active and passive instruments that will be deployed in PLOCAN will be used to fulfil the objective of Task 1 identifying the presence of swim-bladdered fish and the migratory pattern of zooplankton in relation to the sound pressure which they are subjected to. The passive acoustic data that will be acquired exploiting the availability of the W1M3A observatory will be used to achieve the objective of Task 2 about the study of seasonal variability in underwater soundscape, with a specific focus on the noise-wind dependency and impact on the behaviour of marine mammals since the W1M3A is located in the protected area defined as Cetacean's sanctuary.

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

The project will provide sound pressure levels (SPL) that is a quantitative measure of the acoustic energy present in the water column, expressed in decibels relative to a reference pressure of 1 μPa (dB re 1 μPa). It represents the root mean square (RMS) of the sound pressure over a defined time period and is used to describe both continuous and impulsive underwater sounds. In the context of the EU Marine Strategy Framework Directive (MSFD) — specifically Descriptor 11 (Underwater noise and energy) — SPL is one of the key indicators for assessing the impact of anthropogenic noise on the marine environment. Two main types of sound will be considered: Impulsive sound (Indicator 11.1.1): e.g. pile driving, seismic surveys, explosions, characterized by high peak levels and short duration and Continuous low-frequency sound (Indicator 11.2.1): e.g. shipping noise, measured typically as broadband SPL in the 63 Hz and 125 Hz one-third octave bands. The raw acoustic data will be processed to obtain a time-series of SPL that can be delivered in the standard Hierarchical Data Format version 5 (HDF5): it is an open standard developed by *The HDF Group* for efficiently storing and organizing large and complex scientific datasets. It is widely used in marine bioacoustics, oceanography, and environmental monitoring because it allows structured, scalable, and interoperable data storage.

Measurements collected by the Acoustic Zooplankton Fish Profiler (AZFP) will be delivered as a times series of Volume Backscatter (Sv), Target Strength (TS) and will be made at disposal using the Network Common Data Form standard (NetCDF) that is a widely used data format for storing and sharing array-oriented scientific data, particularly in environmental and oceanographic research. The same standard will be used for the oceanographic data (i.e. wind speed, wave height, temperature, salinity) that will be collected as ancillary data.

NetCDF format will be also used for the time-series of echolocation clicks provided by the FPOD.

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

The instruments that will be installed on both PLOCAN and W1M3A sites are autonomous, self-acquiring and powered by batteries, this means that all the data will be stored directly on the memory cards of the instruments themselves and will be directly available to the users upon recovery. Meteorological and oceanographic data acquired by the TA installations will be made available to the users through ERDDAP services already operational in both TA installations.

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).

The raw data will be analysed through routines already developed by the users for quality control based on the Quality Assurance of Real-Time Oceanographic Data (QARTOD) standard that provides a comprehensive manual for real-time quality control of passive acoustic data. It outlines procedures for validating data from hydrophones and other acoustic sensors, ensuring that the collected data meets the necessary standards for scientific analysis. The manual emphasizes the importance of calibration, sensor verification, and data validation to maintain data integrity.

The netCDF files will be OCEANSITES compliant for QC flag and labels.

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

The data can only be processed after the instruments are recovered, as they are autonomous and do not provide real-time data. After the instruments are recovered from the first installation, they must be cleaned and inspected. Only after these procedures will it be possible to recover the data acquired during the first experiment from the internal memory cards and to prepare them for the second deployment (e.g., replacing batteries and spare parts, shipping). During the period of the second installation at sea, the data recovered from the first experiment will be analyzed. The data acquired during the second experiment will be analyzed only after recovery for the second installation, and this may imply some time for embargo. As indicated in the data summary, time series regarding sound pressure levels, the volume backscattering coefficient (Sv), and target strength will be provided; the data volume could reach the order of magnitude of megabytes. Such data could prove useful to researchers in the fields of marine ecology, acoustic oceanography, and conservation science, as well as to universities, research institutes, marine protected area managers, environmental agencies, and policymakers involved in marine environmental protection.

2. Making data FAIR

2.1 Making data findable, including provisions for metadata

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

We confirm our ability to follow the AQUARIUS Data Management approach for metadata. The only challenge may be actual data availability, as the instruments are battery-powered and some battery failures may occur. We may also encounter some issues due to instrument damage or electronics failures, which we will only be able to discover during the recovery, as real-time data will not be available.

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

We plan to use standard procedures, so that all the information could be easily traceable.

2.2. Making data openly accessible

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 research plan includes two experiments lasting approximately six months each, so some time will be needed to process all the acquired acoustic data, especially considering that the raw

acoustic observations, by their nature, consist of a huge number of samples (more than several terabytes). We envisage a one-year embargo for the provision of time series of SPL, Sv, TS, echolocation pulses, and calls, while ancillary oceanographic and environmental observations (e.g., temperature, salinity, wind speed, wave height) will be provided in near-real time and made freely accessible via the Erddap server of PLOCAN and of the W1M3A, already operational, for inclusion in EMODNET and Copernicus INSTAC. The envisaged embargo will be also used to allow scientific analyses preparatory to publish the results of 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?

Oceanographic and environmental observations (e.g., temperature, salinity, wind speed, wave height) will be available for open access download via the Erddap server of PLOCAN and of the W1M3A, already in place and operational. Data products such as marine mammal detections, or the annotated datasets will be included in public repositories, such as ZENODO, to guarantee their traceability through a DOI.

2.3. Making data interoperable

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?

We confirm our ability to gather the input required for the AQUARIUS TA Data Summary Log, and we don't face any challenges in providing the required information.

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

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

We strongly support and believe in the principles of open access, promoting free and unrestricted sharing of scientific data, so there are currently no problems to follow the license CC-BY-4.0

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 research plan does not include the generation of other research outputs than collected data, but the final goal will be the publication of the obtained results.

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)¹ 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.

The contact person will be Sara Pensieri from CNR and her e-mail address is sara.pensieri@cnr.it

5. Data security

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?

As soon as the data is retrieved from each instrument's internal memory cards, a backup of the flash cards will be performed under the responsibility of each instrument owner. Each participating institution will be responsible for providing long-term backups of the data provided by their instruments to network-attached storage (NAS) devices and local storage devices. Other methods for sharing and backing up acquired data, such as a virtual research environment (VRE), will be evaluated if available at the research group's institutions. PLOCAN, as part of the TA team, ensures data integrity and confidentiality through its compliance with the *Esquema Nacional de Seguridad (ENS)* at the Medium Level. This certification guarantees that all data handling, temporary storage, and transfer operations are performed within a secure digital environment, featuring multifactor authentication, continuous network monitoring, and robust backup and recovery procedures. These measures provide a certified framework for data protection before transfer to the AQUARIUS expert data centres.

6. Ethical aspects

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

No issues identified now.

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 data management plan provided by AQUARIUS includes the procedures followed by the team group.

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