



AQUARIUS Data Management Plan AR7E-2026

Jun 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 Data Management Plan – AR7E-2026

1. Data Summary

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

The data collected will

- Be openly available to researchers everywhere following the GO-SHIP principles and data availability policy, with end-of-cruise and updated quantitative data published with a DOI via BODC/SeaDataNet as well as at CCHDO, image data will be made available through EcoTaxa with an assigned DOI, the processed plankton abundance and size distribution data will be published with a linked DOI at BODC and EuroBIS, and genomic sequence data archived at the UK NCBI
- Be used by the user group to study changes in property transports across the boundary of the Subpolar North Atlantic, and the relationships among physical, biogeochemical, and biological variables and how they vary across the region

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

- Quantitative data as continuous time series (underway navigation, meteorology, and surface ocean properties; single-beam/multibeam centre-beam bathymetry), continuous time series of depth profiles (underway upper ocean velocities), discrete time series (surface salinity, isotopes, nutrients, Chl-a and HPLC pigments, and phytoplankton cell abundance), continuous water column profiles at multiple sites (CTD/LADCP data), discrete water column profiles at multiple sites (carbon, nutrients, salinity, oxygen, transient tracers, isotopes, Chl-a and HPLC pigments, and phytoplankton cell abundance), harmonised using NERC controlled vocabularies and community standards for metadata (e.g. [SeaBASS](#) protocols) and combined in NetCDF format files
- Images of individual plankton cells (JPEG format) collected from the underway system and from discrete water column samples will be stored as raw images, along with a metadata file (TSV format) and both will be uploaded and stored in EcoTaxa.
- Quantitative data as plankton abundance and size distribution data processed from above images will be saved in CSV format.
- Genomic data as DNA sequences will be stored in FASTQ sequence file format. Processed taxonomic abundances will be saved in CSV format
- Non-QC'd incidental data: quantitative data as gridded 2-D series in Geotiff format (multibeam bathymetry data), with both acoustic returns and sound speed profile information

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

- Data gathered from ship-supplied sensors (CTD, LADCP, SADC, underway) will be transferred via the shipboard computing systems to user-supplied workstation computers, by users on the ship
- Data gathered from user-supplied sensors and instruments (N₂O/CH₄, plankton images) will be transferred from user-supplied acquisition devices either via the shipboard computing system or via external hard drives (USB drives) to user-supplied workstation computers, by users on the ship
- Samples collected on the ship will either be analysed on the ship (salinity, carbon, nutrients, dissolved oxygen, transient tracers) or will be preserved/prepared and transferred as freight or luggage (depending on the sample type) to users' shore-based laboratories for analysis, and the derived data will be transferred to the user group lead via file sharing services (for bottle sample data) or uploaded directly through the relevant data centre portal

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).

- 1) GO-SHIP parameters: CTD sensor data, Niskin or surface underway bottle sample data, underway sensor data, LADCP and SADC data will be processed according to GO-SHIP standards (Hood et al., 2010; Wilson et al., 2018) for quality assurance, calibration and averaging. The processing will be implemented using the CODAS toolbox (SADC data), the LDEO IX toolbox (LADCP data), and the ocp_hydro_matlab toolbox (CTD, surface underway, and bottle sample data, and interface to LDEO_IX). The ocp_hydro_matlab toolbox will be released and published with Zenodo at the different stages when relevant data are submitted (see next question), and the data file metadata will include the release/version information for each toolbox used as well as documenting processing options (e.g. calibration functions or unit conversions applied). The NERC controlled vocabularies will be used for variable and flag labels and the files will be formatted following CCHDO netCDF file conventions for CTD and bottle data, and as CF-compliant NetCDF-format files for ADCP and underway data. HPLC pigment data processing will be processed following established laboratory protocols for phytoplankton pigment analysis and reported using SeaBASS-compatible field names, units and metadata where applicable. Pigment concentrations will be reported in standard concentration units and harmonised with CTD/bottle metadata. Lugol-preserved phytoplankton samples will be analysed by light microscopy using an inverted microscope and the Utermöhl sedimentation method. Taxonomic references will be standardised against WoRMS following standard community practices, and abundance tables will be formatted in CSV.
- 2) Biological/plankton data: Image data will be processed using the EcoTaxa web application for annotation and classification of the different taxa and the identification of 'non-living' particles, following phytoplankton imaging best practices (Clayton, S. et al., 2026). Processed individual abundance data will be saved as .csv format, using WoRMS taxonomy. Data processing for eDNA metabarcoding will be done using appropriate and specific pipelines and bioinformatic tools. Read quality will be assessed with FastQC (Andrews 2010) and MultiQC (Ewels et al. 2016). Taxonomic assignment will be done using standard classifiers.

When do you plan to perform these processing steps?

- 1) a. CTD, LADCP, SADC and other underway sensor data including CH₄/N₂O, and ship-analysed water sample data (salinity, oxygen, nutrients, carbon, transient tracers) will be processed on board the ship on an ongoing basis, resulting in preliminary processed data at the end of the cruise
- b. Final updates, quality checks and harmonisation of the above will be performed within one month after the cruise (salinity, oxygen and nutrient samples, CTD and underway T, S and O data, ADCP and underway navigation and meteorology data) resulting in a partial final dataset
- c. or within 4 months after the cruise (inorganic carbon and transient tracer sample data, preserved pigment (Chl-a) and DOC samples, CTD and underway fluorescence), resulting in a full final dataset of GO-SHIP hydrography within 6 months of the end of the cruise
- d. with the exception of the following: oxygen isotopes, CH₄/N₂O (samples, and final calibrated sensor data), and radioisotopes we expect to be completed by May 2027, with the delay due to a move of the physical laboratory (radioisotopes) and to the relevant technicians having recently announced plans to leave the lab, resulting in delays while recruiting their replacements (oxygen isotopes and CH₄/N₂O)
- 2) a. Image data will be processed to extract quantitative features within 2-4 months after the cruise, then over the following 4-6 months the over 1 million images will be uploaded to the EcoTaxa web platform and taxonomic annotation and community characterisation will be performed.

b. Lugol-preserved phytoplankton samples will be analysed within 8 months after the cruise and final abundance tables within one month thereafter, and HPLC pigment samples within 12 months after the cruise with QA/QC and harmonisation complete within one month after the analytical results.

c. eDNA: funding is being sought through related projects and pending proposals currently, which would enable analysis in the 12-18 months after the cruise. If someone else is funded to extract and analyse these samples before that, we would collaborate to make the samples available to them.

What is the expected size (Megabyte to Terabyte range) of the data? Please note that in case of multibeam or imagery data, the derived products (such as grids) should be submitted instead of the raw measurements.

- 0) Unprocessed data (gridded multibeam data): 500 MB
- 1) GO-SHIP parameters: 100 MB
- 2) a. Plankton images: 25GB, b. Plankton quantitative data: 10 MB

To whom might your data be useful, outside your project?

These data will be useful to researchers studying ocean circulation, changing heat and freshwater budgets, and physical, biogeochemical and biological contributors to ocean nutrient and carbon cycles in the Subpolar North Atlantic and at larger scales. Data from near Greenland may also be useful to researchers and managers looking to monitor coastal biodiversity. The GO-SHIP data will also contribute to ocean atlases and to calibration/validation datasets for autonomous observing systems including Argo.

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 (physics, chemistry, geophysics, geology, and biology), EuroBIS (marine biodiversity), ELIXIR-ENA (biogenomics), ICOS-Ocean45 (carbon), and Copernicus INSTAC (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 WP5. In addition, as part of 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 or SeaDataNet SeaNoe service for transfer of your metadata and data packages to the AQUARIUS expert data centres. Paragraph 5.4 of Deliverable D6.2 lists the metadata that EMODnet Ingestion requests you to enter at least in order to document your data packages submission. In addition, Annex 1 of Deliverable D6.2 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.

What naming conventions for your data files will you follow?

File names will follow the convention(s)

CE26008_ ITZOOCYGXV_{expocode}_{datatype}_{number or datetime}[_{processing type}[_{revision}]].{suffix}

e.g.

CE26008_ITZOOCYGXV_45CE20260723_ctd_001_2db_0.nc (CTD cast 1 data averaged to 2 dbar, first version),
 CE26008_ITZOOCYGXV_45CE20260723_oxy_all_QC_1.nc (bottle dissolved oxygen data, with flags, revised),

CE26008_ITZOOCYGVX_45CE20260723_PlanktonImage_20260802104430_annotated.jpg (image of one plankton from the specified time, with annotations added),
 CE26008_ITZOOCYGVX_45CE20260723_pigments_HPLC.csv (HPLC pigment data),
 CE26008_ITZOOCYGVX_45CE20260723_phyto_cellabundance.csv (Phytoplankton cell abundance data from microscopy)
 etc.

Plankton image data collected with the IFCB will be organised using the standard IFCB naming conventions where three files are recorded for each individual sample, organised by data, by year, and finally by deployment:

CE26008_ITZOOCYGVX/2026/D20260715/D20260715T000000_IFCB225.hdr (file containing the instrument level metadata for sample D20260715T000000 collected at 00:00:00 on 15/07/2026)

CE26008_ITZOOCYGVX/2026/ D20260715/D20260715T000000_IFCB225.adc (file containing the particle level measurement and metadata)

CE26008_ITZOOCYGVX/2026/ D20260715/D20260715T000000_IFCB225.roi (file containing the raw binary image data for each image in the sample)

These data files are paired with a cruise level metadata file:

CE26008_ITZOOCYGVX_45CE20260723_PlanktonImage_metadata.tsv (file containing all metadata for the IFCB plankton images).

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

We will document the information related to data collection and sample preparation at sea using the Aquarius metadata app and TA summary log.

For quantitative data, the processing methodology and format/label/flag explanations described in the processing toolbox documentation will serve as a starting point, and during processing the handling of outliers/missing data and other specifics will be added to the digital metadata of the data files and ancillary files by the software tools used to perform the processing and QC steps; this information will include a reference to the specific version(s) of the processing toolbox(es) used.

For plankton imaging data the taxonomic characterization will be done by a expert plankton taxonomists and automated ML classification pipelines for which subsets of annotations are validated by an expert. Unidentified species as well as other non-biological images (debris, 'non-living' material) will be noted. We will generate a .csv annotation table that will indicate the taxonomic classification applied to each image, wither it was manually or automatically assigned and by which operator, and the level of confidence if it is an automated classification. Open access to all the images will be available via the EcoTaxa web application. The image data will also be used to derive the size distribution of plankton cells, by estimating cell sizes from the images classified as plankton. Genomic sequence data will be QC'd using the fastQC toolbox, and the data analysed using the Dada2 pipeline (or other appropriate workflow available at the time of analysis). The tables of taxonomic abundances obtained from both the imaging and the metagenomics, as well as the plankton size distribution data, will be exported into .csv format files for documentation and storage.

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.

There is no embargo as such, but a few types of samples are being collected on an opportunistic basis and funding/capacity to analyse them are still to be determined. This includes

- Radionuclide, oxygen isotope, and CH₄/N₂O samples: analysis subject to recruitment of technicians that has just started (oxygen isotopes and CH₄/N₂O) or to moving equipment to a new laboratory in the next few months (radionuclides), so we expect delays to analysis, however, these should be completed and the data QC'd/harmonised (including calibrated CH₄/N₂O sensor time series) before the end of 2027
- genomic samples: funding is being sought at NOC through several routes to perform the extraction and analysis of these samples, but it may not be in place in time to deliver final data by the end of 2027. Therefore, if other researchers do have funding and capacity in place before we do, we would be happy to make the samples available to them to analyse and share the resulting data publicly.

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 data and metadata will also be available from BODC (quantitative data), EcoTaxa (image data), NCBI (genomic data) and OBIS (biodiversity data). The hydrographic (CTD and bottle sample) data will additionally be made available at CCHDO, the GO-SHIP GDAC.

In case of use of exotic data formats for original data, what methods or software tools are needed to read and understand the data and could you provide documentation about that (including possible open-source code)?

No exotic formats are proposed but the data processing toolboxes we will use are open source and details are given in section 1.

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 gives a checklist which documentation TA researchers are recommended to maintain for this purpose.

How are you going to organise gathering the input required for the AQUARIUS TA Data Summary Log during the TA activities?

On each of the two watches, the designated watch lead on the science team will be responsible for using app on the tablet to enter the required information. Paper and/or digital event logger logs may also be kept for data collection/sampling events as redundancy (using the Annex 1 list as a guide). On a daily basis the Chief Scientist and co-Chief Scientist will review the log and investigate any missing information with the relevant watch lead and/or team/parameter leads.

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

If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

N/A

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 ocp_hydro_matlab toolbox used to process CTD/bottle sample and certain underway data will be updated and the version used for processing on this cruise will be published at Zenodo as a software release.

Samples for genomic analysis may be generated as a result if funding and capacity are not in place to fully analyse them before the end of Aquarius.

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)⁴⁹ 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.*

Who will be responsible for data management in your TA project?

The PI/Chief Scientist is responsible for these data management efforts within the project, with the watch leads contributing to using the app (as described above) and parameter PIs contributing to QA/QC and to documentation of processing/standard procedures.

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?

Onboard the ship the quantitative data and metadata will be synced hourly to three user-supplied Linux workstations with hourly backups (including change tracking) to three external hard drives, while the image data will also be backed up to three dedicated external hard drives on a three hourly basis. At the end of the cruise, one hard drive for each dataset will be shipped back and the rest will be hand-carried by different people back to NOC. The preliminary (end-of-cruise) data will then be uploaded to NOC's site-wide backed up filesystem by the user group lead or the relevant team lead (physics team for quantitative data, biology team lead for imaging data), and at this point available image data will be transferred to NOC's secure ftp server.

For quantitative data analysed (e.g. samples shipped back for analysis in a shore lab) or further processed/QC'd after the cruise (e.g. tracer data), the responsible person for each parameter will use email or online file sharing to transfer their finalised data to the user group lead at NOC, who will add it to NOC's backed up filesystem to combine with other parameters measured from the same locations before submitting to the national data centre, BODC, which will make the data available to SeaDataNet. Image data acquired from samples after the cruise will be stored on NOC's ftp server initially then submitted to EcoTaxa. Genomic data will be stored in external hard drives, initially and then submitted to NCBI.

6. Ethical aspects

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

We will adhere to the Nagoya protocols where relevant, and undertake our obligations related to access and benefit sharing of the genetic resources we collect and analyse.

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?

Quantitative data will be submitted to BODC.

References

Andrews, S. (2010) FastQC: A Quality Control Tool for High Throughput Sequence Data. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

Clayton, S., Neeley, A., Poulton, N., et al., (2026) Operational Phytoplankton Observations Best Practices: a guide for using imaging technologies for routine monitoring of phytoplankton communities. Version 1.0.0. Woods Hole, MA, Ocean Carbon and Biogeochemistry Program, Operational Phytoplankton Observations Working Group, 96pp. DOI: <https://doi.org/10.25607/OBP-2059>

Philip Ewels, Måns Magnusson, Sverker Lundin, Max Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, *Bioinformatics*, Volume 32, Issue 19, October 2016, Pages 3047–3048, <https://doi.org/10.1093/bioinformatics/btw354>

Hood, E.M., C.L. Sabine, and B.M. Sloyan, eds. 2010. The GO-SHIP Repeat Hydrography Manual: A Collection of Expert Reports and Guidelines. IOCCP Report Number 14, ICPO Publication Series Number 134. Available online at <http://www.go-ship.org/HydroMan.htm>

Wilson, Samuel & Bange, Hermann & Arévalo-Martínez, Damián & Barnes, J. & Borges, Alberto & Brown, Ian & Bullister, John & Burgos, Macarena & Capelle, David & Casso, Michael & de la Paz, Mercedes & Farias, Laura & Fenwick, Lindsay & Ferrón, Sara & Garcia, Gerardo & Glockzin, Michael & Karl, David & Kock, Annette & Laperriere, Sarah & Rehder, Gregor. (2018). An intercomparison of oceanic methane and nitrous oxide measurements. *Biogeosciences*. 15. 5891-5907. 10.5194/bg-15-5891-2018.