

C-SHIFT

Coastal SST high-resolution inference from shipboard transects

Project funded through AQUARIUS

(Europe Framework Programme for Research and Innovation under Grant Agreement
No 101130915)

Starting date: 1 May 2026

Duration: 12 months

Data Management Plan

2026-06-12

University of Gothenburg

Project coordinator: Pierre De Wit

Infrastructure used: NORSOOP

AQUARIUS initial (**final**) Data Management Plan template for TA Call applications

September 2024

AQUARIUS initial (and final**) Data Management Template - to be completed and submitted as part of the TA Call application**

Project answers in italics

1. Data summary

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

C-SHIFT aims at producing data on sea surface temperature, salinity and spectral EM radiance in the northern Skagerrak through continuous measurements (once/minute) from the Color Hybrid ferry, which makes four trips between Sandefjord, Norway and Strömstad, Sweden every day. These data will be used in a related project which aims at groundtruthing satellite Sea Surface Temperature (SST) data through in situ measurements.

Satellite measurements have a relatively coarse spatial resolution, which prevents accurate measurements of SST near coastlines, especially complex ones such as the Swedish and Norwegian Skagerrak coasts. Therefore, we have developed machine learning methods to increase the resolution. However, the computer-modeled data need to be validated with in situ measurements ranging from near-coast to the open sea. For this, the ferry tracks are ideal. C-SHIFT therefore aims to make the ferry-based observations openly available through the internet, so that they can be used for model validation.

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

The project will generate three types of data from observations from the Color Hybrid ferry, using the NORSOOP infrastructure, which is managed by NIVA in Norway.

Every day of the project period, the ferry schedule is:

Sandefjord – Strömstad 10:00 – 12:30

Strömstad – Sandefjord 13:30 – 16:00

Sandefjord – Strömstad 17:00 – 19:30

Strömstad – Sandefjord 20:00 – 22:30

During these transects, the following measurements are taken:

- *Ferry Box temperature (once/minute) from 2026-05-01 – 2027-04-30 (one year)*
- *Ferry Box salinity (once/minute) from 2026-05-01 – 2027-04-30 (one year)*

Both of these are measured using a SeaBird SBE 45 Thermosalinograph

- *Hyperspectral radiance (once/minute) from 2026-08-25 – 2027-02-24 (six months)*

Measured using a TriOS RAMSES radiometer.

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

Answer

Data is transmitted from each instrument to the local computer and time aggregated. Therefore, the data is transferred via SFTP to the NIVA cloud storage and undergo QC prior to publishing daily in a NetCDF/CF format.

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).

Answer

T and S measurements go through automatic quality control tests and are labelled with quality, but no data processing is applied on AutoQC RT level. Ramses sensor data undergoes quality control that check sun position and ship position in relation to the sun, to avoid problematic measurements due to ship shadow, sun glint or not enough light. Missing Ramses sensor data is kept as a gap, as it is expected that correct measurements will not be possible continuously during a 24-hour period. Published datasets are Climate and Forecast (CF) compliant. Implemented quality tests follow the recommendations in Quality Control of Biogeochemical Measurements DOI <http://doi.org/10.13155/36232>

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

Answer

The QC is applied automatically during ingestion of each file, which takes approximately 15 minutes.

For the final DMP also questions in red:

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.

500 Mb

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

These data will be used in a separate project, funded by the Swedish Foundation for International Cooperation in Research and Higher Education (STINT), in order to ground-truth satellite observations of sea surface temperature. The purpose of this project is to be able to better forecast marine heat waves along complex coastlines by increasing the spatial resolution of satellite data using machine learning algorithms.

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-Ocean** (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 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** and/or **SeaDataNet SeaNoe service** for transfer of your metadata and data packages to the AQUARIUS expert data centres.

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.

What naming conventions for your data files will you follow?

Answer

We follow [ACDD](#) and Climate and Forecast convention

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

*Datasets are documented using **ACDD-compliant metadata**, enabling indexing and discovery through standard catalog services. The metadata is translated into other metadata standards, for example, iso19115. Use of standardized metadata conventions facilitates integration with catalogues (adc.met.no, nordatanet, geonorge.no ++) ensuring the data reaches as wide as possible. Metadata including links to quality-controlled data will be available from the Norwegian Meteorological Institute, as well as other public data portals.*

Metadata for data generated in the project will be logged in the Aquarius app by the project coordinator, as noted in the Aquarius Data Management Plan (<https://zenodo.org/records/13642869>).

Temperature and salinity data are already being submitted by the infrastructure to SeaDataNet through the EMODnet data ingestion pipeline on a regular basis. Hyperspectral irradiance data will be submitted in the same database post-collection and post quality control using the same pipeline.

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

Answer

All the data undergoes the same procedures; processing and QC steps are documented and publicly available on GitHub and in the metadata for the published data.

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.

We will not produce any data that will require a temporary embargo.

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?

*Data are made available via **standard APIs and services**, including protocols such as **OPeNDAP and CSW**, enabling both human and machine access. Data generated in*

the project will be made accessible on the internet by the infrastructure (NORSOOP / NIVA) through the NIVA Thredds Data Portal (<https://data.niva.no/thredds-data-catalog>). Raw data from the ship transects will be updated on a daily basis. Quality-controlled data will be published at the end of the project period.

If possible, there will be a data integration with ferry box data from the Tjörnö and Kristineberg Marine Laboratories, which are published live online as temperature and salinity time series plots through the University of Gothenburg Marine Infrastructure (<https://www.weather.mi.gu.se/en/>). This would ideally continue after the conclusion of the project as well, as the process will be automated.

Quality-controlled datasets generated in this project will be stored centrally on NIVA servers, where they will continue to be accessible for future use. The data will be linked to a DOI, to ensure that they become easily findable as well.

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)?

Answer

NetCDF/CF is a standard format in the field with broad open-source support, for example xarray.

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?

We do not foresee any challenges with maintaining the summary log.

*The use of **NetCDF/CF**, based on the **Common Data Model (CDM)** and controlled vocabularies, ensures interoperability across platforms and disciplines.*

*The format is widely adopted and recognized as an **OGC standard**, and is used by organizations such as WMO and major initiatives like Copernicus and GHRSSST.*

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

As we are only gathering three datasets, the summary log will be quite simple. We will have regular online meetings between the PIs and the TA staff to ensure that the input will be gathered. In addition, we will have a physical 2-day meeting on August 24th-25th to coordinate the data gathering and transfer.

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?

Published data will be made available using Commons Licences (CC BY or CC0) to enable third parties to access, mine, exploit, reproduce and especially disseminate our results.

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 project will provide data for ground truthing machine-learning modelled satellite data, with the goal of observing and forecasting marine heat waves on the spatial scale of individual bays along complex coastlines.

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)** 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 NORSOOP infrastructure, managed by NIVA in Norway, will be responsible for the data management.

*The project leverages **existing community standards (NetCDF/CF, OPeNDAP, CSW)**, minimizing development and maintenance costs. The existing NORSOOP infrastructure is reused to ensure a **cost-effective and sustainable approach** to data management and sharing.*

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?

*Data storage is compatible with **modern storage solutions**, including secure cloud-based object storage systems with recovery and backup included. Write access to the data is handled SSO and two-factor authentication for selected users.*

6. Ethical aspects

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

No specific ethical issues are identified related to the data format or dissemination approach.

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 approach aligns with broader European data ecosystems, enabling interoperability with infrastructures such as **EOSC and Destination Earth**.*