Data Management Plan

MARIS

Dick M.A. Schaap

Funded by
the European Union

About this document

Title	D6.2, Data Management Plan
Work Package	WP6, Data Management & Open Science Practices
Lead Partner	MARIS
Lead Author (Org)	Dick M.A. Schaap (MARIS)
Contributing Author(s)	Hong Minh Le (RBINS), Sissy Iona (HCMR), Susana Diez (CSIC-UTM), Elena Partescano (OGS), Helge Sagen (IMR), Luminita Buga (NIMRD), Lennert Schepers (VLIZ), Mikael Hedblom (SMHI), Eoin O'Grady (MI), Sylvie van Iseghem (IFREMER), Tjerk Krijger (MARIS), Alessandra Giorgetti (OGS)
Reviewers	Nathalie Tonne (SSBE), Frank Armstrong (MI) Bernadette Ní Chonghaile (MI), Aodhán Fitzgerald (MI)
Due Date	31.08.2024, M6
Submission Date	30.08.2024
Version	**1.0**

Dissemination Level

X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Acronyms and abbreviations

NODC   National Oceanographic Data Centre
TA   Transnational Access

Figures

Tables

Table of Contents

1. Introduction

AQUARIUS is a Horizon Europe funded project that runs from 1st March 2024 to 29th of February 2028. The overarching aim is to provide a highly comprehensive suite of integrated research infrastructures appropriate to addressing significant challenges for the long-term sustainability of our unique oceans, seas and freshwater ecosystems. An impressive range of 57 research infrastructure services will be made available by Transnational Access (TA) Calls to include research vessels, mobile marine observation platforms, fixed marine facilities, experimental research facilities, river & basin supersites, aircraft, drones, satellite services, and sophisticated data infrastructures. By collecting new data and generating new information and knowledge through TA projects, AQUARIUS will support the development phase of the EU Mission to Restore our Ocean and waters by 2030, the Sustainable Blue Economy Partnership, the European Green Deal, and international climate initiatives. It will also contribute to the European Digital Twin of the Ocean and the UN Decade for Ocean Sciences.

The intention is to launch two Super Integration AQUARIUS TA Funding Calls inviting European and international researchers to submit proposals. Each Call (one at Month 9 and one at Month 16) will have targeted themes for each of the four lighthouse regions, consisting of Mediterranean Sea, Atlantic Ocean and Arctic Coasts, Baltic and North Sea basins, and Black Sea, along with their associated rivers. Another call [Call 3] may be launched if infrastructures that can be integrated are still available after the first two Calls, and there is sufficient need, demand and capacity. Proposals will be evaluated by a Scientific Review Panel, especially on the basis of scientific and technical quality. Each Call will be open for specified periods; ten weeks and eight weeks respectively, for the submission of proposals. Experience has shown that the time from the closing of a Call to the start of the logistical evaluation is around five months, depending on the number of applications received and the availability of external reviewers. This implies that TA projects might be deployed for the 1st Call from M16 onwards and for the 2nd Call from Month 23 onwards, while the overall AQUARIUS project runs for 48 months.

As part of the 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 Partnership targets and associated initiatives and projects with the generation of 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, such as SeaDataNet[1] (physics, bathymetry, chemistry, biology, geology), EurOBIS[2] (biodiversity), ELIXIR-ENA[3] (biogenomics), ICOS-Ocean[4] (carbon), and Copernicus INSTAC[5] (Near- Real-Time data), for quality assurance, long term stewardship, and wide access and use. These infrastructures in turn are feeding into EMODnet, Copernicus Marine, Blue-Cloud (EOSC), Digital Twin of the Ocean (DTO) developments, and globally to e.g. GEOSS, and the digital ocean ecosystem that is being developed in the framework of the UN-IOC Ocean Decade programme.

---

[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

This Deliverable D6.2 is dedicated to the overall AQUARIUS Data Management Plan (DMP), and it will describe the AQUARIUS Data Management approach as planned. The AQUARIUS approach will cover the elements as included in the Horizon Europe DMP template.

2. AQUARIUS Data Management approach

2.1. AQUARIUS TA Research Infrastructure Services

AQUARIUS is providing a highly comprehensive suite of integrated research infrastructures (RIs) appropriate to addressing significant challenges for the long-term sustainability of our unique oceans, seas and freshwater ecosystems. An impressive range of 57 research infrastructure services will be made available by Transnational Access (TA) Calls to include research vessels, mobile marine observation platforms, aircraft, drones, satellites, sensors, fixed freshwater and marine observatories and test sites, experimental facilities, and sophisticated data infrastructures. The following figure gives an overview of these Research Infrastructure services.
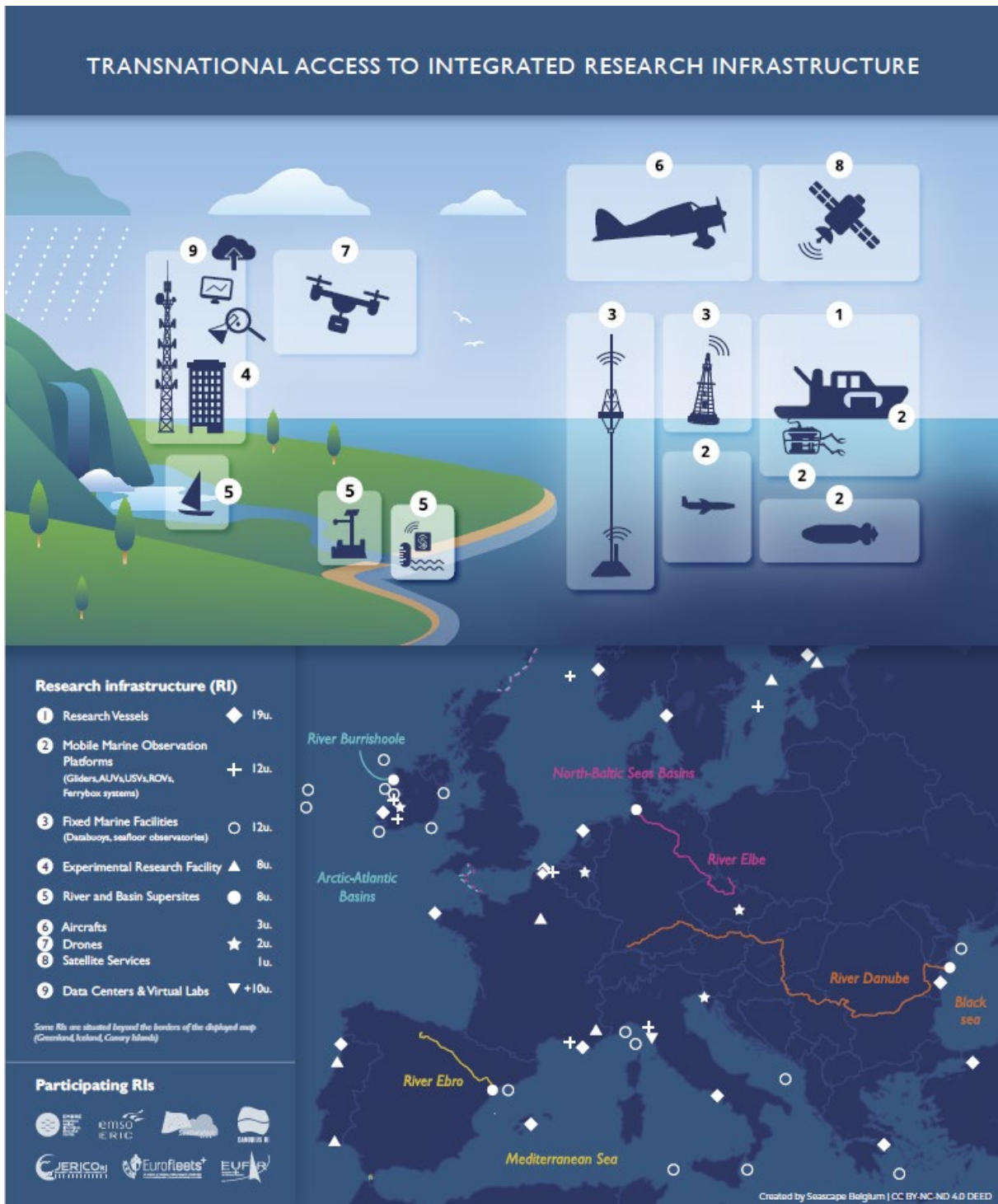
Figure 2.1 Overview of the types of research infrastructures offered by AQUARIUS, and the marine and freshwater areas within the scope of the Lighthouse regions.

The 57 different infrastructures are grouped according to 13 different infrastructure types, for instance: 1) Research Vessels; 2) Mobile Marine Observation Platforms; 3) Fixed Marine Facilities; 4) Experimental Research Facility; 5) River and Basin Supersites; 6) Aircraft; 7) Drones; 8) Satellite Services; and 9) Data Centres & Virtual Labs. AQUARIUS thereby brings together leading RIs in Europe, to include EMBRC, EMSO-ERIC, SeaDataNet, DANUBIUS-RI, JERICO RI, Eurofleets+, and EUFAR.

The aim is to provide free-of-charge access to these Research Infrastructure services through the AQUARIUS Transnational Access Calls, whereby proposers will be requested to combine multiple services in their proposals.

To learn more about the diverse Research Infrastructure services, a survey has been held by the Marine Institute, gathering information about the services regarding their capabilities, modes of operation, and data management practices, where available. The results of the survey will be used as part of the TA Calls for proposals to inform interested researchers about the Research Infrastructure services that will be made available for TA activities.

The replies to the data management questions are less elaborated and quite diverse. As the Calls for TA proposals will aim at combining various Research Infrastructure services, AQUARIUS will not focus on trying to harmonise the data management practices of the various operators of the Research Infrastructure services. However, it will focus on the TA scientific teams adopting a common AQUARIUS approach for managing and publishing the resulting data as collected during TA projects and data products and knowledge as generated as part of the TA projects. This is a more feasible and practical approach to secure a maximum return of investment of the AQUARIUS TA activities for contributing to the EU Mission 'Restore our Ocean and Waters by 2030', the Sustainable Blue Economy Partnership, the European Green Deal, the European Digital Twin of the Ocean, and international initiatives such as the UN Decade for Ocean Sciences.

2.2.   European marine and ocean data management landscape

Oceanographic, marine and freshwater data include a very wide range of measurements and variables, originating from a broad, multidisciplinary spectrum of projects and programs. Several thousands of research institutes, governmental organisations, and private companies collect oceanographic, marine and freshwater data. Various heterogeneous observing sensors are installed on research vessels, submarines, aircraft, and moorings, drifting buoys, gliders, floats, fixed platforms, and satellites. These sensors measure physical, chemical, biological, geological and geophysical parameters, with further data resulting from the analysis of water and sediment samples for a wide variety of parameters. The AQUARIUS RI portfolio offers many of these instruments and facilities for TA scientific teams to collect new data sets and generate new data products and scientific insights.

It is of great importance that maximum benefit can be derived from data once it has been acquired. The principles of "capture once – use many times" and achieving 'FAIRness'[6] are major targets for managing and serving the wealth of ocean, marine, and freshwater data sets to the existing and potential user community. Large amounts of multidisciplinary data need to be made available in an interoperable or harmonised and standardised way in order to carry out cross-domain analysis and processing of data that is necessary for gaining better understanding of the ocean and its functioning.

---

[6]https://www.nature.com/articles/sdata201618

Since the early nineties a wide range of EU initiatives, funded and/or supported by ECDG RTD[7], EC DG MARE[8], EC DG DEFIS[9], EC DG ENV[10] and EC DG CONNECT[11] have been undertaken to develop a European capacity for collecting and managing marine in-situ and remote sensing data, while federating and interacting with national activities for developing data centres and data management systems, and contributing to international and global initiatives.

Great progress has been made with developing standards, services, and establishing dedicated European infrastructures for the management of marine and freshwater data. These infrastructures provide services for discovery and access to collected data and for ensuring long-term stewardship. These concern long-term marine data initiatives such as EMODnet[12] (bathymetry, chemistry, geology, physics, biology, seabed habitats, and human activities), and Copernicus Marine[13] (CMEMS: ocean analysis and forecasting), mandated by the European Commission to deliver open access marine data and data products, and several well established, leading, European marine data management research infrastructures, such as SeaDataNet[14] (physics, chemistry, geophysics, bathymetry, geology, and biology), EurOBIS[15] (marine biodiversity), Euro-Argo[16] (ocean physics and marine biogeochemistry), ELIXIR-ENA[17] (biogenomics), and ICOS-Ocean[18] (carbon).

Each of these data management research infrastructures has established links to data originators and their data collection, facilitating overseeing and engaging in the process from collection to validation to storage and distribution. This is organised through managing networks of national data centres and expert organisations that are also connected at national level to other actors. Together the national nodes develop and operate the European infrastructures, formulating and sharing common standards and associated tools and services. These European data management research infrastructures are seen as European data aggregators, bringing together and harmonising major volumes of data and metadata from the many data originators. They function as pillars under EMODnet and Copernicus Marine, feeding them with harmonised metadata and data collections on an operational basis, which gives major input for the generation of European and global data products by EMODnet and Copernicus Marine. Several of the European data management research infrastructures are also themselves increasingly involved in generating data products and models, which are run by the infrastructure teams or made available as services for external users from research, government and industry. These blue data infrastructures are also mostly complementary, dealing with other data originators and/or different stages in the processing chains from data acquisition to data products to knowledge.

---

[7]EC DG RTD: Research and Innovation

[8]EC DG MARE: Maritime Affairs and Fisheries

[9]EC DG DEFIS: Defence Industry, and Space (formerly known as EU DG GROW)

[10]EC DG ENV: Environment

[11]EC DG CONNECT: Communications Networks, Content and Technology

[12]https://emodnet.ec.europa.eu/en

[13]https://marine.copernicus.eu/

[14]https://www.seadatanet.org

[15]https://www.eurobis.org

[16]https://www.euro-argo.eu/

[17]https://www.ebi.ac.uk/ena/

[18]https://otc.icos-cp.eu/
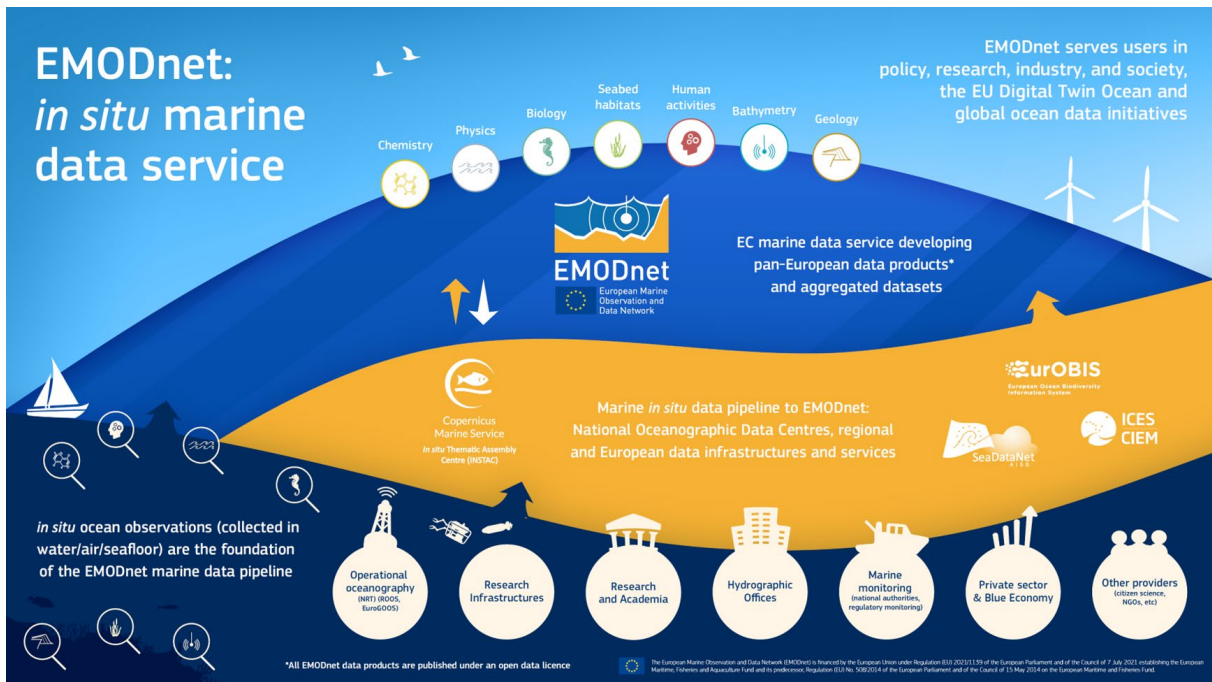
Figure 2.2 Illustration from EMODnet of the European data pyramid in which data management research infrastructures are organising the inflow of data from many sources to EMODnet for generating derived data products serving a wide user community (source: EMODnet[19])

Moreover, they are working together in many EU projects relevant for the marine domain, but also around the European Open Science Cloud[20] (EOSC). Technological advances provide innovative opportunities for new forms of science, which is one of the drivers behind EOSC.

However, this demands well-described, accessible data that conforms to community standards. The FAIR principles articulate the attributes data need to have to enable and enhance reuse, by humans and machines. EMODnet is making use and integrating output from several of these data management infrastructures to generate and give open and free access to standardised and harmonised data collections and derived data products, based upon data from hundreds of data producers and providers. For analyses, various research groups actively develop and run algorithms, and explore cloud-based science. A considerable achievement is the development of the Copernicus Marine Service. This service provides regular and systematic reference information on the physical state, variability and dynamics of the global ocean and the European regional seas.

Another major achievement is the development and deployment of the Blue-Cloud[21] infrastructure, which provides a federated Data Discovery & Access service for many leading marine repositories, and a Virtual Research Environment (VRE), hosting multiple Virtual Labs and offering a variety of services. This infrastructure has proven its abilities for supporting solving complex challenges, combining data and algorithms, in a cloud-based manner in the Blue-Cloud H2020 pilot project, and the HE Blue-Cloud 2026 project which has succeeded since January 2023. Blue-Cloud is also considered as the marine thematic EOSC.

---

[19]https://emodnet.ec.europa.eu/en

[20]https://open-science-cloud.ec.europa.eu/

[21]https://www.blue-cloud.org

Moreover, there is the Destination Earth[22] (DestinE) programme shaping Europe's digital future, that aims to develop a platform to generate and utilise Digital Twins of the Earth. In the meantime, several projects have been launched for developing Digital Twins of the Ocean (DTO) in projects such as ILIAD and EDITO. It is clear that EMODnet, Copernicus, Blue-Cloud and RIs will play a major role in achieving the challenges and promises of the DTO.

2.3. Targets for AQUARIUS data management

As part of the 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).

In order to ensure quality assurance, long-term stewardship and wide access and use of the data collected in the AQUARIUS TA projects, it is envisaged that the data and metadata, and possible resulting data products, become part of the repositories managed and operated by the leading European data management infrastructures, namely SeaDataNet23 (physics, bathymetry, chemistry, biology, geology), EurOBIS24 (biodiversity), ELIXIR-ENA25 (biogenomics), ICOS-Ocean26 (carbon), and Copernicus INSTAC27 (Near- Real-Time data). Achieving this, will also ensure that the AQUARIUS data and data products will become available for EMODnet, Copernicus Marine, Blue-Cloud (EOSC), Digital Twin of the Ocean (DTO) developments, and globally to e.g. GEOSS, and the digital ocean ecosystem that is being developed in the framework of the UN-IOC Ocean Decade programme.

As a second target, AQUARIUS wants to introduce and educate researchers of the TA project scientific teams in the opportunities of open science as promoted and encouraged by EOSC. For that purpose, AQUARIUS will have a synergetic cooperation with the Blue-Cloud 2026 project, facilitating TA researchers to explore and make use of the Blue-Cloud Virtual Research Environment (VRE) and related applications for further analyses and processing of AQUARIUS data into new data products and knowledge, expressed in potential scientific publications. Such additional output should also become available in a FAIR way, using the research infrastructure channels, Blue-Cloud catalogue, EOSC and Zenodo[28], the EU Open Research Repository. This could also include new analytical workflows and/or software that might be developed by AQUARIUS TA researchers in the Blue-Cloud 2026 e-infrastructure.

---

[22]https://digital-strategy.ec.europa.eu/en/policies/destination-earth

[23]https://www.seadatanet.org

[24]https://www.eurobis.org

[25]https://www.ebi.ac.uk/ena/

[26]https://www.icos-cp.eu/observations/ocean/otc

[27]https://marine.copernicus.eu/about/producers/insitu-tac

[28]https://zenodo.org/

3. AQUARIUS Expert Data Centres and VRE experts

The goals of the AQUARIUS data management approach are making data and data products resulting from TA activities openly available in line with FAIR principles through established European data management infrastructures and within the lifetime of the AQUARIUS project. The primary responsibility for achieving this lies with the TA scientific teams that will commit themselves by means of the contracts that they will sign with AQUARIUS as part of the Call award procedure.

However, to get a maximum result and impact, the TA scientific teams will be supported and work together with expert data centres. These are partners in the AQUARIUS project, experienced in marine data management, and well acquainted with the principles, standards, and related tools and services for implementing the AQUARIUS data management approach. Most of them are National Oceanographic Data Centres (NODCs) and well connected to the European data management infrastructures that are promoted in AQUARIUS for achieving open access following FAIR principles. In addition, they are geographically spread over Europe, present in and well informed about the characteristics of the four EU Mission Lighthouse regions, and together share expertise about the different data disciplines of relevance for AQUARIUS.

The expert data centres are listed in the following table.

Organisation	Data management infrastructure(s) related to	Lighthouse region(s)	Involved in AQUARIUS WP?
CSIC - Spain	NODC, SeaDataNet, EurOBIS, EMODnet	Mediterranean, Atlantic-Arctic	WP5, WP6
IFREMER - France	NODC, SeaDataNet (operator), Copernicus INSTAC (operator), EurOBIS, EMODnet	Mediterranean, Atlantic-Arctic	WP5, WP6
RBINS - Belgium	NODC, SeaDataNet, EMODnet	Baltic-North Sea	WP5, WP6
IMR - Norway	NODC, SeaDataNet, EurOBIS, Copernicus INSTAC, EMODnet	Atlantic-Arctic	WP5, WP6
VLIZ - Belgium	NODC, SeaDataNet, EurOBIS (operator), EMODnet	Baltic-North Sea	WP5
EMSO-ERIC - Italy	EMSO-ERIC	Mediterranean, Atlantic-Arctic, Danube - Black Sea	WP5, WP6
MARIS - Netherlands	SeaDataNet (operator), EMODnet	Mediterranean, Atlantic-Arctic, Danube - Black Sea, Baltic-North Sea	WP5, WP6

Organisation	Data management infrastructure(s) related to	Lighthouse region(s)	Involved in AQUARIUS WP?
OGS - Italy	NODC, SeaDataNet, EurOBIS, Copernicus INSTAC, EMODnet	Mediterranean	WP5
HCMR – Greece	NODC, SeaDataNet, EurOBIS, Copernicus INSTAC, EMODnet	Mediterranean, Danube - Black Sea	WP5, WP6
EMBL-EBI	ELIXIR, BioSamples, ENA, MGnify, PRIDE, Metabolights, BioImageArchive	Global	WP5-WP6
EMBRC-ERIC - France	EurOBIS, SeaDataNet	Mediterranean, Atlantic-Arctic, Danube - Black Sea, Baltic-North Sea	WP5, WP6
SMHI - Sweden	NODC, SeaDataNet, Copernicus INSTAC, EMODnet	Baltic-North Sea, Atlantic-Arctic	WP5, WP6
INCDM - Romania	NODC, SeaDataNet, EMODnet	Danube - Black Sea	WP5, WP6
NORCE - Norway	ICOS-Marine (operator)	Mediterranean, Atlantic-Arctic, Danube - Black Sea, Baltic-North Sea	WP5, WP6
MI - Ireland	NODC, SeaDataNet, EurOBIS, Copernicus INSTAC, EMODnet	Baltic-North Sea, Atlantic-Arctic	WP6
SSBE - Belgium	EMODnet	Mediterranean, Atlantic-Arctic, Danube - Black Sea, Baltic-North Sea	WP6

Table 3.1: Overview of AQUARIUS expert data centres

These expert data centres will work together with the TA scientific teams from the moment that the TA proposals have been awarded in order to make the TA scientific teams more aware and educated in the AQUARIUS data management approach and to support them in the actual deployment before, during, and after the TA activities. They also will work together with the organisers of the AQUARIUS Calls for TA proposals, giving information about the AQUARIUS Data Management approach. This will be done by publishing the Deliverable D6.2 (this report), giving presentations at AQUARIUS brokerage events and answering questions of potential TA call applicants. Moreover, they will give support for reviewing the initial Data Management Plans (DMP phase 1) following the AQUARIUS template (see paragraph 5.2) that TA call applicants will be asked to complete and deliver as part of their TA proposal. For practical reasons, expert data centres will form DM teams per lighthouse region.

The cooperation between the TA scientific teams and the expert data centres will be done as part of the Work Packages 5 and 6, in particular in:

- <u>Task 5.5: Data Management and data stewardship Training</u> – M1-M30 [Lead: HCMR, with CSIC, IFREMER, RBINS, IMR, VLIZ, EMSO-ERIC, MARIS, OGS, EMBRC-ERIC, SMHI, INCDM, NORCE]
  This training will inform and educate researchers on European marine data infrastructures, FAIR standards, services and tools to use for processing, documenting, and ingesting their newly collected marine data and resulting data products. The training programme will be compiled as a series of presentations and hands-on exercises and will be deployed through a number of online workshops. The expert data centres will act as trainers. The webinar(s) will be recorded and made available on the AQUARIUS website for outreach and Call promotion.

- <u>Task 6.2: FAIR Data Management of data to be collected and processed during and after TA projects</u> - M1-M48 [Lead: MARIS, with CSIC, EMBRC-ERIC, EMSO-ERIC, NORCE, IFREMER, HCMR, INCDM, MI, SMHI, RIBNS, IMR and SSBE]
  This task involves a number of preparatory activities, which should be completed before the launch of the TA Calls respectively before the awarded TA projects are going to be implemented. These include:
  - o Formulating a federated and FAIR Data Management flow scheme, which is done in this Deliverable D6.2. This Deliverable will be made available as part of the information and guidelines that will be provided on the TA Portal (TAP) for potential TA proposers.
  - o Developing and deploying the federated data flow scheme system in the coming period before the start of the first TA projects (planned from M16 onwards).
  - o Implementing the AQUARIUS Data Management approach together with the TA scientific teams from M16 onwards, following the federated data flow scheme and using its components, that will be specified in this Deliverable D6.2.

Related Deliverables will be:

- D6.2: Data Management Plan – MARIS – M6 (Plan for a federated Data Management flow scheme and integrated AQUARIUS Dataflow Dashboard (ADD))

- D6.3: Operational ADD – Phase 1 and 2 – MARIS – M12 (Operational federated and integrated AQUARIUS Dataflow Dashboard (ADD) with selected functionality 1 (M12) and 2 (M24))

- D5.4: Report on Data Management and data stewardship Training – HCMR – M24

- D6.4: TA Data Management follow-up – I & II – MARIS – M24 (Data management follow-up report for funded TA events I (M24) and II M48))

In addition to the AQUARIUS data management activities, there is also an objective in AQUARIUS to introduce the researchers from the TA scientific teams in the principles and practices of open science as promoted by the European Open Science Cloud (EOSC) initiative of the EU. This will be implemented through synergy between the HE AQUARIUS project and Blue-Cloud 2026 project.

The Blue-Cloud pilot as part of 'The Future of Seas and Oceans Flagship Initiative' of EU H2020 combined interests of developing a thematic marine EOSC cloud and serving the Blue Economy, Marine Environment and Marine Knowledge agendas. It deployed a cyber platform with a smart federation of multidisciplinary data repositories, analytical tools, and computing facilities in support of exploring and demonstrating the potential of cloud based open science for ocean sustainability. Since early 2023, the successor HE Blue-Cloud 2026

project works at a further evolution into a Federated European Ecosystem to deliver FAIR & Open data and analytical services, instrumental for deepening research of oceans, EU seas, coastal & inland waters. Blue-Cloud 2026 is a core project as part of the Mission Ocean. Its key services are a federated Blue-Cloud Data Discovery & Access service (DD&AS), a versatile Blue-Cloud Virtual Research Environment (VRE), supported by e-infrastructures such as D4Science, EUDAT, and EGI, and an increasing portfolio of Blue-Cloud Virtual Labs.

As part of AQUARIUS, TA scientific teams will be invited and encouraged to register and make use of the Blue-Cloud Virtual Research Environment (VRE), which hosts a range of analytical software services, several Virtual Labs, and a structured framework for setting up and running new VLabs as analytical pipelines. It also features services for FAIR documenting (provenance metadata), visualising, and publishing of the analytical data products in a product catalogue which is shared with the European Open Science Cloud (EOSC) and Zenodo.

The researchers of the TA scientific teams will be introduced to, and will experience, web-based open science. It will allow them to elaborate their new data sets by deeper analyses and by combining these with existing data and data products from a range of leading European data repositories and data product generators. This will facilitate further research to better understand and manage the many aspects of ocean sustainability.

This will be implemented in a cooperation of TA scientific teams with AQUARIUS partners that are experts in the Blue-Cloud Virtual Research Environment (VRE) and Virtual Lab applications.

The Blue-Cloud VRE experts are listed in the following table:

Organisation	VRE / VLabs	Involved in AQUARIUS WP?
CNR - Italy	Blue-Cloud VRE (operator) and Blue-Cloud 2026 scientific coordinator	WP5, WP6
MARIS - Netherlands	Blue-Cloud DD&AS (operator) and Blue-Cloud 2026 technical coordinator	WP5, WP6
VLIZ - Belgium	Vlabs on Carbon-Plankton Dynamics (leader) and Zoo and Phytoplankton EOV products (leader); EDITO-Infra (co-leader)	WP5, WP6
INGV – Italy	WorkBench on Physics (temperature & salinity) EOVs (leader)	WP5, WP6
OGS - Italy	WorkBench on Eutrophication (chlorophyll, nutrients, oxygen) EOVs (leader)	WP5, WP6
EMBL-EBI – United Kingdom	WorkBench on Ecosystem-level EOVs (co-developer)	WP5, WP6
IH – Portugal	VLab on Integration of coastal ocean observations along Europe (leader)	WP5, WP6

Organisation	VRE / VLabs	Involved in AQUARIUS WP?
AWI – Germany	WebODV service (operator)	WP5, WP6
ULIEGE - Belgium,	VLab on Coastal currents from observations (leader)	WP5, WP6
IRD – France	VLab on Global Fisheries Atlas (leader)	WP5, WP6
MOI - France	Copernicus Marine (coordinator), EDITO-Infra (co-leader), EDITO-Modellab (leader), Blue-Cloud-DTO Taskforce leader	WP6

Table 3.2: Overview of AQUARIUS VRE experts

The cooperation between the TA scientific teams and the Blue-Cloud VRE experts will again be done as part of the Work Packages 5 and 6, in:

● Task 5.6 Virtual Access and Analytics Training – M1-M36 [Lead: CNR, with: VLIZ, INGV, OGS, EMBL, IH, AWI, ULIEGE, and IRD]
This training will prepare and educate the TA researchers in the EOSC strategy and web-based open science practices. They will learn about the Blue-Cloud Virtual Research Environment and its analytical applications, that they can use for further elaborating their newly collected data and combining these with large existing European and global data collections for deepening their scientific insights and knowledge. The training programme will be compiled as a series of presentations and hands-on exercises and will be deployed through a number of online workshops. The Blue-Cloud VRE experts will act as trainers. The webinar(s) will be recorded and made available on the AQUARIUS website for outreach and Call promotion.

● Task 6.3: Applying open science by analysing and valorising collected TA project data in a web-based research environment - M18-M48 [Lead: MARIS, with CNR, VLIZ, MOI, INGV, OGS, EMBL-EBI, IHPT, AWI, ULiege and IRD]
This task involves several activities, aimed at researchers from TA scientific teams trying out the Blue-Cloud VRE and applications for generating added-value data products, which could be published as AQUARIUS data products. These include:
  o Registering TA project scientific teams in the Blue-Cloud Virtual Research Environment (VRE); This interacts with Task 5.6 activities for educating and training of TA project scientific teams in web-based open science.
  o Facilitating and coaching TA project scientific teams by Blue-Cloud experts to make use of the analytical services, rich federated marine data repositories for in-situ and satellite data as well as data products, and work with the Virtual Labs at the Blue-Cloud VRE, for elaborating their newly gathered data, and combining these with other data sets, into new data products, information, and knowledge insights, relevant for the Mission and Partnership;
  o Blue-Cloud experts collaborating with TA project scientific teams for documenting and publishing resulting data products and publications in EOSC and the AQUARIUS Dataflow Dashboard (ADD).
  o The Blue-Cloud VRE will be provided as VA service by CNR (Italy).

Related Deliverables will be:

- D5.4: Report on Data Management and data stewardship Training – HCMR – M24

- D6.5: VA Data Analytics follow-up - I & II – MARIS – M36 (Data analytics follow-up report for funded VA access to Blue-Cloud VRE related to TA events I (M36) and II (M48))

4. Main datasets gathered and managed in AQUARIUS considering FAIR principles

One major aim of the AQUARIUS project is to gather new environmental data sets and derived data products and knowledge for the lighthouse regions in support of the EU Mission to Restore our Ocean and waters by 2030, the Sustainable Blue Economy Partnership, the European Green Deal, European Digital Twin of the Ocean developments, and international initiatives, such as the UN Decade for Ocean Sciences. This will be done by gathering and generating new data sets for various scientific disciplines through the AQUARIUS TA projects which will be undertaken by teams of researchers making use of combinations of the 57 AQUARIUS research infrastructure services that are offered for TA exploitation.

To establish a maximum return of investment, AQUARIUS has an open data policy, obliging TA scientific teams to deliver and publish their new data sets and derived data products within the lifetime of the AQUARIUS project and using well established European data management infrastructures for optimal public discovery and access.

Another major aim of the AQUARIUS project is to inform and educate researchers from the TA scientific teams in the opportunities and benefits of open science by facilitating them to make use of the Blue-Cloud VRE and applications for further analysing their new collected data sets and possibly combining these with data from large repositories to derive added-value data products and scientific knowledge. Also, these data products and knowledge should be published and made available for discovery and access through public services, such as the Blue-Cloud data products catalogue, which is shared with EOSC, while resulting publications are shared through Zenodo.

The FAIR concept relates to "Data and services that should be Findable, Accessible, Interoperable, and Re-usable, both for machines and for people." The emphasis is on machine FAIRness. In the following paragraphs, the current status for the Blue-Cloud will be described, and reference will be made to activities which are already planned as part of the Blue-Cloud Work Plan and to activities which might be added, for improving FAIRness. The management of the AQUARIUS data and data products should be done following the FAIR principles which are integrated in the AQUARIUS Data Management approach.

4.1. Making data findable, including provisions for metadata

Data are Findable when they are described by sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications. The identifier enables persistent linkages to be established between the data, metadata and other related materials in order to assist data discovery and reuse. Related materials may include the code or models necessary to use the data, research literature that provides further insights into the creation and interpretation of the data and other related information.

For AQUARIUS, use will be made of leading European marine data management infrastructures, namely SeaDataNet (physics, bathymetry, chemistry, biology, geology), EurOBIS (biodiversity), ELIXIR-ENA (biogenomics), ICOS-Ocean (carbon), and Copernicus INSTAC (Near- Real-Time data), for quality assurance, long term stewardship, and wide access and use. These infrastructures in turn are feeding into EMODnet, Copernicus Marine, Blue-Cloud (EOSC), Digital Twin of the Ocean (DTO) developments, and globally to e.g. GEOSS, and the digital ocean ecosystem that is being developed in the framework of the UN-IOC Ocean Decade programme. Each of these preferred infrastructures has developed and is operating dedicated discovery and access services, applying community standards and principles. Moreover, they are well engaged in activities for analysing and improving

their FAIRness, for instance as part of the ENVRI-FAIR, FAIR-EASE, and ENVRI-HUB-NEXT projects. While, for the results of open science, the Blue-Cloud VRE data products catalogue will give users overview and access to all data products generated in the VRE and its Virtual Labs, also following common standards.

The infrastructures to be used by AQUARIUS for data management are accessible both for humans and machines through well-defined and universally implementable protocols. Anyone should be able to access at least the metadata. During embargo periods, to be agreed with AQUARIUS scientific teams, selected data sets might be temporarily restricted.

4.2. Making data interoperable

Interoperable data and metadata should be described by community and/or domain standards for technical interoperability and vocabularies for semantic interoperability, and they include qualified references to other data or metadata. It is this that will allow the data to be 'machine-actionable'. Interoperability is an essential feature in the value and usability of data. Legal interoperability of data must be considered as well. In FAIR, legal interoperability falls under the principle that data should be 'Reusable'. It is planned that the delivered data in AQUARIUS will follow the formats as prescribed by the preferred European marine data management infrastructures. The infrastructures also make use of specific vocabularies for marking up metadata and data. Information about the formats, vocabularies and appropriate software and services can be found at the portals of these infrastructures. But in AQUARIUS, training and coaching activities will be undertaken by expert data centres and Blue-Cloud VRE experts to inform, educate, and support the TA scientific teams in adopting the prescribed metadata and data formats and QA-QC for uptake of the resulting AQUARIUS data in those leading infrastructures.

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

For data to be Reusable, the FAIR principles reassert the need for rich metadata and documentation that meet relevant community standards and provide information about provenance, reporting how data was created and information about consecutive data reduction or transformation processes to make data more usable, understandable or 'science-ready'. The ability of humans and machines to assess and select data on the basis of criteria relating to provenance information is essential to data reuse, especially at scale. Reusability also requires that the data be released with a 'clear and accessible data usage licence': in other words, the conditions under which the data can be used should be transparent to both humans and machines. In the AQUARIUS framework the CC-BY-4.0 will be adopted as default licence as it gives a very low threshold, and it is supported and promoted by the preferred European infrastructures and the Blue-Cloud VRE data products catalogue. The latter catalogue aims at publishing data products as generated by the researchers with the Virtual Labs. Their metadata is enriched with the provenance documentation, this way providing users detailed information about the way that the data products were generated, which will be very useful for wider use. And the provenance information gives scientific users detailed information how the data products might be reproduced with the Virtual Labs which is again very useful as it allows VRE users to repeat the same analyses, but for instance changing input and/or settings.

5. AQUARIUS Data Management flow scheme

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.

Steps	Activities	Support
Step 1a – pre-planning	• **AQUARIUS Data Management Plan D6.2** included in guidelines for TA proposers	Expert data centres can provide advice upon request
Step 1b – pre-planning	• TA project proposers complete initial Data Management Plan for their project, following **AQUARIUS DMP template phase 1**, and include this in their TA project proposal	Expert data centres review DMP phase 1 in support of Call evaluation process
Step 2a – planning	• After awarding of TA project, TA proposers are asked to refine their initial Data Management Plan for their project, following **AQUARIUS DMP template phase 2** • **Validated TA DMP phase 2** will be published on **AQUARIUS Dataflow Dashboard (ADD)**	Expert data centres give support and advice to TA project scientific team MARIS
Step 2b – planning	• Preparation of **promotional factsheet of the TA project** and publishing on AQUARIUS website • **Link to promotional factsheet** of TA project included in AQUARIUS Dataflow Dashboard (ADD)	WP7 team together with TA project scientific team MARIS
Step 3 – training in data management	• **Training in data management** organised by webinars for TA project scientific teams • Recorded webinars and presentations made available on AQUARIUS website	Expert data centres, led by HCMR WP7 team
Step 4 - deployment	• **Data gathering and processing** undertaken by TA project scientific teams, supported by providers of AQUARIUS research infrastructure services • During TA activities, PI of TA project scientific team is requested to maintain an **AQUARIUS TA Data Summary Log** with an index of data gathering activities • In case the TA project will involve scientific cruises with Research Vessels, then also a **Cruise Summary Report**	Expert data centres are available for advice and support for data management issues MARIS provides a Tablet with App for maintaining TA Data Summary Log

Steps	Activities	Support
	(CSR)[29] should be completed by the TA project PI within a month after the cruise.	TA project scientific team (PI)
	● **Completed TA Data Summary Log** for selected TA project will be published on AQUARIUS Dataflow Dashboard (ADD)	MARIS
	● If applicable, a **link to the Cruise Summary Report** on the SeaDataNet portal will be published on the AQUARIUS Dataflow Dashboard (ADD)	MARIS
Step 5 – data processing and transfer	● TA project scientific teams undertake efforts for **transforming metadata and data to standard formats** as valid for applicable European data management infrastructure(s)	Coaching and support by expert data centres
	● Transfer of **elaborated metadata and data as packages**, sorted for data types / disciplines, through the **Data Submission service of EMODnet Ingestion**[30] **or SeaDataNet SeaNoe service**[31]**,** to expert data centres	Reception of data and metadata packages and assignment to selected expert data centres
	● **Links to submissions in EMODnet Ingestion and SeaNoe** will be published on AQUARIUS Dataflow Dashboard (ADD)	MARIS
Step 6 – publishing of data and metadata in FAIR way	● **Received data and metadata** will be reviewed by assigned expert data centres with feedback to TA project scientific teams.	Expert data centres in contact with TA PIs
	● **Validated data and metadata** will be included in local data centres and populated into the applicable **European data management infrastructure(s)**	Expert data centres
	● **Links to inclusions in European data management infrastructure(s)** will be published on AQUARIUS Dataflow Dashboard (ADD)	MARIS
Step 7 – training in open science practices	● **Training in open science** and use of the Blue-Cloud VRE and its applications,	Blue-Cloud VRE experts, led by CNR

---

[29]https://www.seadatanet.org/Metadata/CSR-Cruises

[30]https://www.emodnet-ingestion.eu

[31]https://www.seadatanet.org/Software/SEANOE

Steps	Activities	Support
	organised by webinars for TA project scientific teams ● Recorded webinars and presentations made available at AQUARIUS website	WP7 team
Step 8 – practising open science	● TA researchers encouraged to register in Blue-Cloud VRE	CNR
	● TA researchers **performing analytics** on their newly acquired data sets, possibly in combination with data from established data repositories, using Blue-Cloud applications	Coaching and support by Blue-Cloud VRE experts
Step 9a – publishing of data products	● TA researchers documenting **open science data products** with metadata in the Blue-Cloud VRE data products catalogue which are shared with EOSC and Zenodo	Coaching and support by Blue-Cloud VRE experts
	● **Links to published data products** will be published on AQUARIUS Dataflow Dashboard (ADD)	MARIS
Step 9b – publishing of scientific publications (papers)	● TA researchers preparing potential **scientific publications (papers)** about their TA projects which could be published with DOIs in Zenodo	TA researchers
	● **Links to scientific publications** will be published on AQUARIUS Dataflow Dashboard (ADD)	MARIS
Step 10 – FAIR indicators	● Expert data centres maintaining and publishing **best practices and FAIRness indicators** for the AQUARIUS data flow and its achievements; inclusion by VRE experts about uptake and results of open science practices	Expert data centres and VRE experts

Table 5.1: Overview of the AQUARIUS Data Management Flow scheme

Note: the TA projects might deal with different types of data that could require splitting of packages over multiple expert data centres. In addition, some data types will be available directly during the TA activities (Real time mode), while others are samples that need to be processed in laboratories and/or with dedicated software, which will take extra time and could also imply waiting times (Delayed mode). Therefore, the expert data centres together will use the AQUARIUS TA Data Summary Log as a 'shopping list' to know what data is coming, who to contact, and how to divide over expert data centres. They will also maintain a follow-up scheme so that there is a clear overview of assignments, data flow status, and achieved throughput.

The following paragraphs will detail the various components to be used in the sketched data management flow scheme.

5.1. Promotional factsheet of TA project

As part of the promotion activities in WP7, factsheets will be produced with well-illustrated narratives about all scheduled TA projects with information to be gathered from the TA project scientific teams. These factsheets will be published by WP7 on the AQUARIUS website. Links to each TA project factsheet will then be included in the AQUARIUS Dataflow Dashboard, which will be a platform in the AQUARIUS website to allow users to follow the data gathering, processing, and publishing of AQUARIUS data, data products and scientific publications for each TA project. The information in the AQUARIUS Dataflow Dashboard will be publicly available.

5.2. AQUARIUS Data Management Plan template for TA projects

Each TA proposal applicant for a funded TA project must complete an initial Data Management Plan (DMP) and include it as part of the TA proposal. 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. The AQUARIUS expert data centres will assist with the Call process and respond to questions about the AQUARIUS data management approach, for as far as answers cannot be found in this Deliverable D6.2 which will be published on the TA Platform (TAP) and AQUARIUS website as guidance. 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 expert data centres will form DM Teams for each of the EU Mission Lighthouse regions. During the kick-off they will explain the overall approach and goals for Data Management in AQUARIUS, give background on the preferred European data management infrastructures for selected data types and disciplines, their guidelines, standards, and associated tools and services, and the expected results of the Data Management activities. Thereafter, they will go together with the TA teams through the initial DMP for refining existing answers and completing new questions.

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. The following paragraphs give the templates of the TA project DMP phase 1 and phase 2.

5.2.1. AQUARIUS: TA project DMP Template - Phase 1: Preliminary DMP

(to be filled in as part of the TA proposal submission)

1. Data Summary

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

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

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

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).

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

2. Making data FAIR

2.1 Making data findable, including provisions for metadata

The **AQUARIUS Data Management approach**[32] is aimed at inclusion of your collected/generated data in leading European aquatic data management infrastructures such as **SeaDataNet**[33] (physics, chemistry, geophysics, geology, and biology), **EurOBIS**[34] (marine biodiversity), **ELIXIR-ENA**[35] (biogenomics), **ICOS-Ocean**[36] (carbon), and **Copernicus INSTAC**[37] (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**[38] and/or **SeaDataNet SeaNoe service**[39] for transfer of your metadata and data packages to the AQUARIUS expert data centres.

Paragraph 5.4 of Deliverable D6.2 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 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.

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

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

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.

---

[32] Once D6.2 has been published at Zenodo, a direct link should be included in the DMP template

[33] https://www.seadatanet.org

[34] https://www.eurobis.org

[35] https://www.ebi.ac.uk/ena/

[36] https://www.icos-cp.eu/observations/ocean/otc

[37] https://marine.copernicus.eu/about/producers/insitu-tac

[38] https://www.emodnet-ingestion.eu

[39] https://www.seadatanet.org/Software/SEANOE

Will you collect/generate data that might require a temporary embargo? Indicate which type of data and motivate its temporary embargo for each case.

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?

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.

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?

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?

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?

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)**[40] 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

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?

6. Ethical aspects

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

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?

5.2.2. AQUARIUS: DMP Template - Phase 2: Full DMP

(to be filled in after having been awarded for a TA project, updating and refining the earlier Phase 1 submission. In some cases, a few new questions have been added)

1. Data Summary

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

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

---

[40]https://www.seadatanet.org/Metadata/CSR-Cruises

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

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, data formatting, conforming to standards).

When do you plan to perform these processing steps?

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.

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

2. Making data FAIR

2.1. Making data findable, including provisions for metadata

The **AQUARIUS Data Management approach**[41] is aimed at inclusion of your collected/generated data in leading European aquatic data management infrastructures such as **SeaDataNet**[42] (physics, chemistry, geophysics, geology, and biology), **EurOBIS**[43] (marine biodiversity), **ELIXIR-ENA**[44] (biogenomics), **ICOS-Ocean**[45] (carbon), and **Copernicus INSTAC**[46] (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**[47] or **SeaDataNet SeaNoe service**[48] 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?

---

[41] Once D6.2 has been published at Zenodo, a direct link should be included in the DMP template

[42] https://www.seadatanet.org

[43] https://www.eurobis.org

[44] https://www.ebi.ac.uk/ena/

[45] https://www.icos-cp.eu/observations/ocean/otc

[46] https://marine.copernicus.eu/about/producers/insitu-tac

[47] https://www.emodnet-ingestion.eu

[48] https://www.seadatanet.org/Software/SEANOE

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

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.

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?

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)?

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?

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?

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.

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?

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?

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?

6. Ethical aspects

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

---

[49]https://www.seadatanet.org/Metadata/CSR-Cruises

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?

5.3.   AQUARIUS TA Data Summary Log and dedicated app

The **AQUARIUS TA Data Summary Log** 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 expert data centres as it will provide an index to the collected/generated data and relevant documentation, such as who, when, where, how, etc. It will serve for them as a list to know what data to expect from where and who and as a checklist for the following steps in the data management workflow. The **AQUARIUS TA Data Summary Log** will contain only metadata and no data.

MARIS will develop an app for entering, browsing, and exporting the information of the TA Data Summary Log. The app will then be delivered as a stand-alone application and preloaded on a tablet, that can also work in offline mode and conditions. Each TA project team will be given a tablet in loan for the duration of the TA data acquisition event as this will prevent any issues with non-compliance of app and tablet.

Currently, there is still discussion ongoing in the WP6 group about the attributes for the TA Data Summary Log as it should be feasible to maintain and complete the summary during the TA data acquisition activities while the summary should also give sufficient information for the expert data centres to get good insights in the acquired data sets and to keep track of their processing flow from as collected to fully validated, elaborated and published. It would be useful that the AQUARIUS TA Data summary compiles a list of all the generated datasets with at least: the datasets names (e.g. CTD, ADCP, bottle, nutrients, eDNA...) and the discipline (e.g. Biochemical, physical, geophysical, ...). An extra challenge is the fact that the TA projects will combine multiple TA installations and platforms, possibly operated and provided by multiple research infrastructures. The final attributes will be determined early September 2024, still in time for developing the data model and the app application.

Remark: In case the TA project will involve scientific cruises with Research Vessels, then also a **Cruise Summary Report (CSR)**[50] should be completed by the TA project PI within a month after the cruise. Cruise Summary Reports are the usual means for reporting on cruises or field experiments at sea. Traditionally, it is the Chief Scientist's obligation to submit a CSR not later than two weeks after the cruise. This provides a first level inventory of measurements and samples collected at sea. Currently, the Cruise Summary Reports directory, maintained by SeaDataNet together with its members, covers cruises from 1873 till today from more than 1.500 research vessels: a total of nearly 65.000 cruises, in all European waters and global oceans. This also includes historic CSRs from European countries that have been loaded from the ICES database from 1960 onwards.

5.4.   Metadata of your dataset packages when submitting via EMODnet Ingestion

The following metadata elements should be completed online when submitting your data and metadata packages via the online Submission service of EMODnet Data Ingestion:

---

[50]https://www.seadatanet.org/Metadata/CSR-Cruises

- Title: clearly mention 'AQUARIUS, the TA project identifier and the name of your project
- Abstract
- Keywords of the observation types
- File format types
- Provenance and lineage:
  ◦ Data quality processing: handling of outliers, missing and suspect values and null observations) and data harmonisation (code, label and QC flag explanations, consistent use of headers, data formatting, conforming to standards).
  ◦ Summary of processing methodology: Different basic steps how the data was derived from the device and how it was combined, transformed and processed (including software used) to the stage presented now.
- Dataset documentation: any relevant files. Please attach the DMP.
- Creation date of the dataset
- Author of the dataset: the submitter is automatically recorded as the responsible of the dataset and is considered the author.
  ◦ Name
  ◦ Email address
  ◦ Country
  ◦ Organisation
- Data holding organisation: country, name, email
- Project or programme: please select AQUARIUS from the dropdown list (EDMERP)
- Sea areas
- Coordinate Reference System (CRS): WGS 84, ETRS89,...
- Spatial Extent (Bounding Box): The spatial coordinates of the widest geographical extent: (westBoundLongitude, eastBoundLongitude, southBoundLatitude, northBoundLatitude)
- Temporal Extent: start and stop date of the collection points
- Taxonomic coverage: Report the taxa (use appropriate levels like families, orders or classes) in the abstract.
- Conditions for data sharing: license.
- Embargo: please upload the DMP pdf, so that information on embargoes is known.
- Files: upload one zipped archive

Note: these metadata are requested to describe the data package that you will upload. The data package itself will contain metadata and the actual data, whereby it is recommended to undertake efforts already to transform the metadata and data formats to the standards and vocabularies of the European data management infrastructure which will be targeted for inclusion and publishing of your data and metadata. This will require that you document several aspects of the data acquisition and following processing, so that you can elaborate your data and metadata later from that documentation. Annex 1 gives a checklist which documentation is recommended to maintain.

5.5. AQUARIUS Dataflow Dashboard (ADD)

The AQUARIUS Dataflow Dashboard (ADD) will be developed by MARIS as a dedicated platform, integrated in the AQUARIUS website, to follow progress of the data management flow scheme from planning stage through to publishing of results for each awarded TA project. It should contain up-to-date and public information about TA projects and their progress. The ultimate goal is to give discovery and public access to research data sets as collected and processed and data products as generated by the TA research teams as part of the AQUARIUS TA projects. The following table gives a draft overview of the planned information attributes for the ADD.

Attribute	Links
AQUARIUS TA Project ID	Unique ID per AQUARIUS TA project
TA Project Information	Dedicated link to retrieve a narrative about the scheduled TA project. The narrative should include at least the TA project objectives, who are involved, which lighthouse regions, which scientific disciplines, and which RI facilities are going to be used.
Date start	Start date of the TA data acquisition activities
Date end	End date of the TA data acquisition activities
DMP phase 2	Completed Data Management Plan Phase for the scheduled TA project
TA Data Summary Log	Summary index to the collected/generated data and relevant documentation, such as who, when, where, how, etc. from the TA data acquisition activities
Cruise Summary Report	Dedicated link to retrieve details from SeaDataNet Cruise Summary Reports database, in case of TA cruise with research vessel
Research data package published in EMODnet Ingestion	Dedicated link(s) to retrieve details from metadata and data sets as submitted by TA Teams and published in the EMODnet Ingestion service
Research data package in SeaDataNet SEANOE	Dedicated link(s) to retrieve details from metadata and data sets as submitted by TA Teams in the SeaDataNet SEANOE Data Publishing service
Elaborated research data as published in European data management infrastructures	Dedicated link(s) to retrieve standardised metadata and data from TA projects in SeaDataNet, EurOBIS, ELIXIR-ENA, ICOS-Marine, Copernicus INSTAC services for data discovery and access
Elaborated open science data products as published in Zenodo	Dedicated link(s) to retrieve standardised metadata and open science data products as generated through Blue-Cloud from TA projects in Zenodo for discovery and access

Table 5.2: Overview of the draft contents for the AQUARIUS Data Flow Dashboard

5.6. AQUARIUS Data FAIRness Indicators

FAIR-related monitoring indicator	Description *(Descriptions for FAIR, nurtured by FORCE11 and published in Nature Scientific Data in 2016[51] are reported in italic)*	Achieved in AQUARIUS project by awarded TA team when
1.1 Data is complete	Data deposit is complete: the announced number of collected datasets are completely transferred to Data Centers or published on target repositories	Number of datasets and volume announced in DMP and TA Data Summary Log match with deposited ones. If numbers and/or volumes do not match, discrepancies should be justified by the TA team. Collected datasets are completely transferred or published on EMODnet DI, SEANOE, targeted European data management infrastructures or alternative repository for exceptional dataset
2.1. Data is findable	*F1. (meta)data are assigned a globally unique and persistent identifier* *F2. data are described with rich metadata (defined by R1 below)* *F3. metadata clearly and explicitly include the identifier of the data they describe* *F4. (meta)data are registered or indexed in a searchable resource*	Metadata and data are published on EMODnet DI, SEANOE, targeted European data management infrastructures or alternative repository for exceptional dataset or alternative repository for exceptional dataset which meet F. principles
2.2. Data is openly accessible	*A1. (meta)data are retrievable by their identifier using a standardised communications protocol* *A1.1 the protocol is open, free, and universally implementable* *A1.2 the protocol allows for an authentication and authorization procedure, where necessary* *A2. metadata are accessible, even when the data are no longer available* Data comply with Open research data principle underpinning scientific research results that has no restrictions on its access, enabling anyone to access it.	Data is published with no restrictions and no embargo on EMODnet DI, SEANOE, targeted European data management infrastructures or alternative repository for exceptional dataset which meet A. principles. A moratorium is not a restrictive access 'per se' but as data is temporarily not accessible during the embargo, data under embargo are in AQUARIUS considered not open/accessible until the embargo's end.

---

[51] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

2.3. Data is interoperable	*I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation* *I2. (meta)data use vocabularies that follow FAIR principles* *I3. (meta)data include qualified references to other (meta)data*	Data is harmonised and published on targeted European data management infrastructures which meet I. principles.
2.4. Data is reusable	*R1. meta(data) are richly described with a plurality of accurate and relevant attributes* *R1.1. (meta)data are released with a clear and accessible data usage licence* *R1.2. (meta)data are associated with detailed provenance* *R1.3. (meta)data meet domain-relevant community standards*	Data is harmonised and published under an open licence on targeted European data management infrastructures which meet R. principles. A moratorium is not a restrictive access 'per se' but as data is temporarily not accessible during the embargo, data under embargo is considered here not reusable until the embargo's end.

Table 5.3: AQUARIUS FAIRness implementation indicators

This Table and indicators will be used to monitor the progress and success of the AQUARIUS data management approach. The proportion of actual FAIRness should be calculated on the fraction of the total datasets on which actual data interoperability is achievable. Depending on the data types, not all the collected data will be able to be ingested in the AQUARIUS reference FAIR data repositories.

5.7. Handling exceptional data sets

The AQUARIUS reference FAIR data repositories are:
- SeaDataNet Seanoe service for all kinds of marine environmental data sets
- EMODnet Data Ingestion service for all kinds of marine environmental data sets
- SeaDataNet CDI Data Discovery and Access service for physics, bathymetry, chemistry, biology, geology data sets
- EurOBIS service for biodiversity data sets
- ELIXIR-ENA service for biogenomics data sets
- ICOS-Marine service for carbon data sets
- Copernicus INSTAC service for Near- Real-Time data sets

Exceptional data sets could be data sets of exotic data types which cannot be ingested into AQUARIUS reference data repositories because of the large data file size exceeding the uploading capacities of the reference data repositories (e.g. pictures, satellite images). The exceptional datasets should, in order to be accounted for in the AQUARIUS data FAIRness indicators, be published on an interoperable data repository connected to the European Open Science Cloud, such as e.g. Zenodo and Ecotaxa.

6. Conclusions

This Deliverable D6.2 gives a practical plan for managing the various data types and data sets that will be collected and processed as part of the TA projects that will be awarded and deployed during the AQUARIUS project.

The AQUARIUS Data Management approach aims at inclusion of the collected/generated data in leading European aquatic data management infrastructures such as SeaDataNet[52] (physics, chemistry, bathymetry, geophysics, geology, and biology), EurOBIS[53] (marine biodiversity), ELIXIR-ENA[54] (biogenomics), ICOS-Ocean[55] (carbon), and Copernicus INSTAC[56] (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 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.

Moreover, the AQUARIUS Data Management approach will arrange that data products and scientific publications, generated during the AQUARIUS project as part of the open science practices of TA project researchers, facilitated by the synergy of AQUARIUS with Blue-Cloud 2026, will also be managed and published in a FAIR way using Zenodo[57], the EU Open Science Repository, which is operated by CERN.

The first deployment of TA projects is planned for Month 16, and this should be sufficient time for further detailing and implementing the tools and services that are planned for following and supporting the data flow management process. This concerns developing and operating the AQUARIUS Dataflow Dashboard as a platform integrated in the AQUARIUS website and the app for maintaining a TA Data Summary Log by TA PIs. These developments are underway.

---

[52] https://www.seadatanet.org

[53] https://www.eurobis.org

[54] https://www.ebi.ac.uk/ena/

[55] https://www.icos-cp.eu/observations/ocean/otc

[56] https://marine.copernicus.eu/about/producers/insitu-tac

[57] https://www.zenodo.org

7. ANNEX 1: Checklist for information to document about data acquisition and processing during TA projects

The following is a maximal set of meta-information elements that are relevant for one individual observation or measurement. Elements in **bold** are required if applicable. Please note that only the elements in *italic* are considered the 'value', but for proper interoperability and reusability, all elements are needed.

- **Provide the TA installation platform and instrument used for data collection or sampling**
- **Campaign code as used by AQUARIUS**
- A description of the different features of interest (e.g. sea water column at depth range x, sediment layer at depth range y, benthos communities of gravel beds,…) of each individual measurement.
- Sampling in the case when an in-situ measurement is performed on the feature of interest:
    - **The sampling location coordinates and their Coordinate Reference System**
    - The name of the location
    - Which lighthouse region(s)
    - **In case of a station, the station code**
    - **The time, sampling depths and locations at the beginning of the sampling, and if applicable also at the end**
    - Sea bottom depth at sampling location if relevant
    - Vertical datum: depth reference system
    - The time, length, swath and width over which sampling took place
    - The sampling device: **type, model,** characteristics, calibration information,...
    - **Subsamples** for chemistry, biology or geology: the extent of the subsample, from which part/depth/organ it was taken,…
- The observed property you have observed or measured:
    - The scientific purpose of the measurement or observation operation
    - The ***parameter***
    - **Statistical modifiers** (time-averaged, percentile, standard error,...)
    - The ***time*** *at which the measurement was taken*
    - The **time reference** at which the measurement was taken (UTC, time offset to UTC)
- The procedure you applied to take the measurement:
    - Any **preparatory steps**: sieving, filtration, mixing,,…
    - An indication on the **fraction** (matrix) or combined sample these steps resulted in (e.g. dissolved, particles (180-300um), wet weight/dry weight,…)
    - For chemistry, a description of all **analytical steps** taken
    - The algorithm you have applied to the raw device output if applicable
    - The measurement device: **type**, model, characteristics, calibration information,...
    - The **time** (plus time reference) at which the result of the procedure was known (if it is different from the time at which the measurement was taken)
- The result of the measurement:
    - The ***value***
    - *The **unit** of the value, expressed in units that are considered a community standard*

- Any information on accuracy and **quality control flags** (plus explanation of their meaning)