www.nautilos-h2020.eu

Organisation: ETT SpA

NAUTILOS

D8.3 Data Management Workflow

Version 1.1

Date: 14th September 2021

Doc. Version: V2.2

10.5281/zenodo.7211798

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101000825 (NAUTILOS). This output reflects only the author's view and the European Union cannot be held responsible for any use that may be made of the information contained therein.

Document Control Information

Settings	Value
Deliverable Title	Data Management Workflow
Work Package Title	Data Management
Deliverable number	D8.3
Description	A wide international community is interested in NAUTILOS targeted measurement, hence whenever possible, the developed data management infrastructure, tools and services should allow an easy data flow towards existing infrastructures and integrators globally accepted and used by the ocean observing community. To this end, the data management infrastructure should be designed to serve both internal project' need and international scientific ocean science community and other stakeholders in the field. Data policy should be open and free and state-of-art and international standards and workflows for interoperability should be applied. NAUTILOS data management infrastructure will deliver data and information to a human accessible interface, i.e., an on-line web user interface that will provide features to discover, access, retrieve sensors and platforms data. This web interface would also represent the entry point for all the users (including citizen scientists) with an interest in validated environmental data collections. This document presents and describe the current design and implementation of the NAUTILOS data management infrastructure.
Lead Beneficiary	ETT SpA
Lead Authors	Antonio Novellino
Contributors	Federica Colombo
Submitted by	Federica Colombo
Doc. Version (Revision number)	2.2
Sensitivity (Security):	Public
Date:	14/09/2021
DOI	10.5281/zenodo.7211798

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Action	Date
Gabriele Pieri	Coordinator / WP1 Leader	*Review*	
Victoria Geraskova	Project Manager	*Review*	
Marco Tampucci	For Coordinator / WP8 Leader	*Approve*	
Victoria Geraskova	Project Manager	*Approve*	

Document history:

The Document Author is authorized to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner.

Changes to this document are summarized in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes
V2.0	14-09-2021	Antonio Novellino	Second full draft
V1.1	31-08-2021	Antonio Novellino	Revised first draft
V1.0	19-08-2021	Antonio Novellino	First full draft
V0.0	30-07-2021	Antonio Novellino	Initial draft

Configuration Management: Document Location

The latest version of this controlled document is stored in <location>.

Nature of the deliverable		
R	Report	**X**
DEC	Websites, patents, filing, etc.	
DEM	Demonstrator	
O	Other	

Dissemination level		
PU	Public	**X**
CO	Confidential, only for members of the consortium (including the Commission Services)	

ACKNOWLEDGEMENT

This report forms part of the deliverables from the NAUTILOS project which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101000825. The Community is not responsible for any use that might be made of the content of this publication.

NAUTILOS - New Approach to Underwater Technologies for Innovative, Low-cost Ocean observation is an H2020 project funded under the Future of Seas and Oceans Flagship Initiative, coordinated by the National Research Council of Italy (CNR, Consiglio Nazionale delle Ricerche). It brings together a group of 21 entities from 11 European countries with multidisciplinary expertise ranging from ocean instrumentation development and integration, ocean sensing and sampling instrumentation, data processing, modelling and control, operational oceanography and biology and ecosystems and biogeochemistry such, water and climate change science, technological marine applications and research infrastructures.

NAUTILOS will fill-in marine observation and modelling gaps for chemical, biological and deep ocean physics variables through the development of a new generation of cost-effective sensors and samplers, the integration of the aforementioned technologies within observing platforms and their deployment in large-scale demonstrations in European seas. The fundamental aim of the project will be to complement and expand current European observation tools and services, to obtain a collection of data at a much higher spatial resolution, temporal regularity and length than currently available at the European scale, and to further enable and democratise the monitoring of the marine environment to both traditional and non-traditional data users.

NAUTILOS is one of two projects included in the EU´s efforts to support of the European Strategy for Plastics in a Circular Economy by supporting the demonstration of new and innovative technologies to measure the Essential Ocean Variables (EOV).

More information on the project can be found at: http://www.nautilos-project.eu.

TABLE OF CONTENTS

List of Figures

List of Tables

EXECUTIVE SUMMARY

NAUTILOS project will deploy a set of sensors and samplers to measure a series of environmental variables and descriptors essential to understand the state of the ocean, its dynamics and properties, to quantify the forcing of the atmosphere-ocean boundary and to understand the role the oceans play in Earth's climate.

These variables consists of 14 biogeochemical, biological and ecosystem essential ocean variables (EOVs), *i.e.* inorganic carbon, stable carbon isotopes, dissolved oxygen, inorganic macro nutrients, suspended particulate, ocean colour, ocean sound, phytoplankton biomass and diversity, zooplankton biomass and diversity, turtles, marine birds, marine mammals abundance and distribution, live coral, sea grass cover, microbial biomass and diversity and invertebrate abundance and distribution, two deep ocean observing system (DOOS) specific EOVs, i.e. litter including micro-plastics, seafloor sponge habitat cover and eight MSFD descriptors.

NAUTILOS will then produce and consume a huge amount of heterogeneous data that may be divided into two categories: 1) internal project data, i.e., data outputs from the project itself; and 2) externally valuable environmental data, i.e., real-world application data, that have to be made readily (i.e. as soon as possible) and freely available to European Marine Data Infrastructure, as well as international ocean science community and stakeholders.

NAUTILOS data management infrastructure workflow deals with basic data handling, harmonization and sharing, hence data management workflow deals with data format, metadata and data service.

The document presents the background and the state-of-art for the implementation of an up-to-date data management workflow, then presents the adopted solutions (standards for metadata, tools for data publication, etc.) and examples of how to interoperate with the deployed infrastructure.

BACKGROUND

1.1 Introduction

NAUTILOS[1] is one of two *TechOceanS*[2] projects under the *BG-07-2019-2020 – The Future of Seas and Oceans Flagship Initiative: Part [C] 2020 – Technologies for observations.* It has the aim of developing and testing new technological solutions that will lower the costs of acquiring, deploying and maintaining monitoring and observing stations to fill the *in situ* observational gaps of current ocean observation systems.

For this purpose, NAUTILOS works on filling in existing marine observations and modelling gaps through the development of a new generation of cost-effective sensors and samplers for physical (salinity, temperature), chemical (inorganic carbon, nutrients, oxygen), and biological (phytoplankton, zooplankton, marine macrofauna) essential ocean variables (EOV), in addition to micro- and nano-plastics, to improve our understanding of environmental change and anthropogenic impacts.

Newly developed marine technologies are going to be integrated in different observing platforms and deployed through the use of novel approaches from shore to deep-sea deployments. NAUTILOS is expected to collect and process a huge amount of heterogeneous data. Therefore, the development of a state-of-art, robust and flexible data management infrastructure is a fundamental requisite for project success.

Since a wide international community is interested in NAUTILOS targeted measurements, the developed data management infrastructure, tools and services should allow an easy data flow towards existing infrastructures and integrators globally accepted and used by the ocean observing community.

Therefore, the data management infrastructure should be designed to serve internal project needs, the wider international scientific ocean science community, and other relevant stakeholders. Data policy should be open and free and state-of-art and international standards and workflows for interoperability should be applied.

In addition, the project also proposes to acquire and manage new data (e.g., digital images, micro-plastic observation etc.) whose harmonised data flow has yet to be designed and adopted at international level. The development will follow models and schemes from the already existing infrastructures and proposes itself as the baseline for the establishment of these new data flows. One key outcome of this project is thus to help share more and higher-quality data for a wider set of parameters.

NAUTILOS data management infrastructure will deliver data and information to a human accessible interface, i.e., an on-line web user interface that will provide features to discover, access, retrieve sensors and platform data. This web interface would also represent the entry point for all the users (including citizen scientists) with an interest in validated environmental data collections.

This document presents and describes the current design and implementation of the NAUTILOS data management infrastructure.

[1] https://cordis.europa.eu/project/id/101000825
[2] https://cordis.europa.eu/project/id/101000858

1.2. Definitions

- DATA - oceanographic observational data, data derived from it and, if applicable, gridded analyzed data fields.
- METADATA – “data about data" describing the content, methods, quality, condition, and other characteristics of the data
- PRODUCT INTEROPERABILITY - The ability of computer systems or software to exchange and make use of data and information
- DATA POLICY – set of rules that regulates the production, use and dissemination of data
- OPEN DATA - Open data is data that anyone can access, use and share. An open data policy can also help encourage informed reuse of third-party data.
- FREE and UNRESTRICTED - denotes –non-discriminatory and without charge.
- FREE of CHARGE - in this context: the data are free to any user without any charge for both data and data products
- FAIR data – Findable, accessible, interoperable, and reusable

1.3. Data policy and disclaimer

Data and metadata generated within NAUTILOS and stored at the originating organisation (*i.e.,* the data owners and data providers and responsible entities for data, metadata and quality), are the same that are stored and made available by the NAUTILOS data portal (data assembly centre) and are the same that are available within the data integrator portals and initiatives (*e.g.,* EMODnet, CMEMS, etc).

The NAUTILOS partners are not responsible for any use and misuse made by end-users.

2. Data Management Specifications and Recommendations

2.1. Data value chain

High-quality ocean data and information are crucial for society to underpin economic activities at sea, to understand the oceans' complex role in the planet system, to understand human interactions and impacts, and empower decision makers with evidence-based information for sustainable management of the ocean and adaptation to climate change.

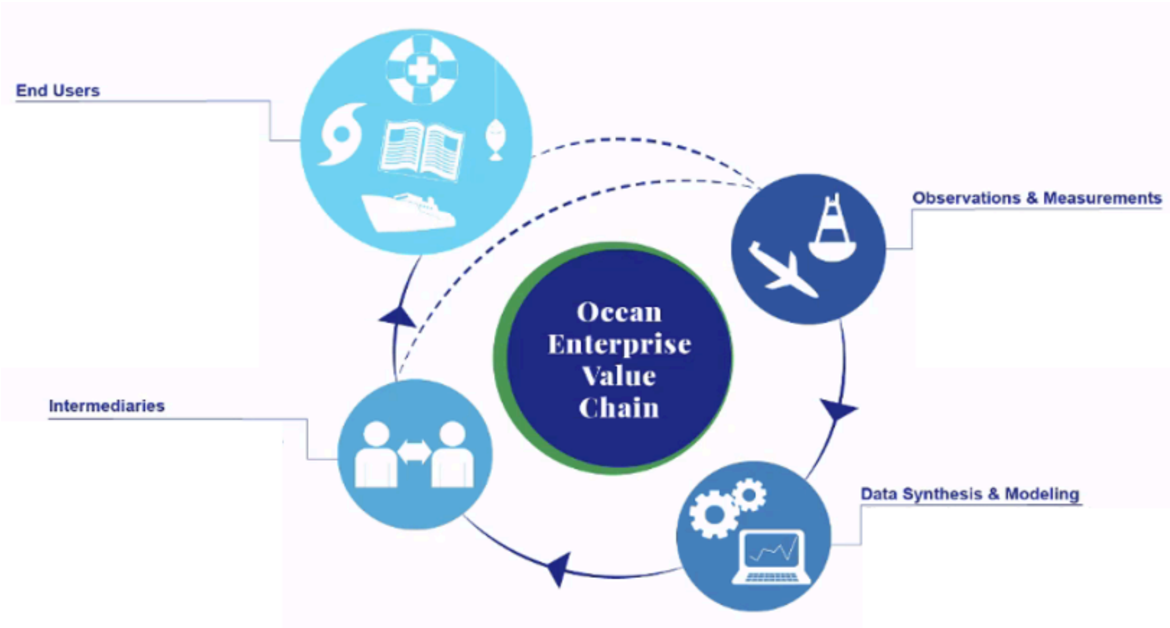

Figure 1. Ocean Observing value chain (adapted from IOOS)

Figure 1 shows the ocean observation data value chain. It is an end-to-end loop in which observations and measurements (O&M) are routinely (*e.g.*, daily) gathered by sensors, data synthesis and modelling integrates into data and data products the collected O&M, intermediaries (governmental, public and private organisations) use these products to create value-added and fit-for-purpose products and end users use these value-added products to make critical safety and economic decisions and plan future actions (e.g., deploy more sensors, more accurate sensors, etc.).

In this loop, the collection of marine observational data has evolved considerably through the advancement of sensor development, autonomous platforms (*e.g.,* moorings, gliders, surface drifting buoys, coordinated Argo profiling floats), remote sensing, etc., and increasing use of ships of opportunity equipped with sensors.

This has further accelerated our understanding of the physical dynamics of the ocean and their interactions with biogeochemistry and biology.

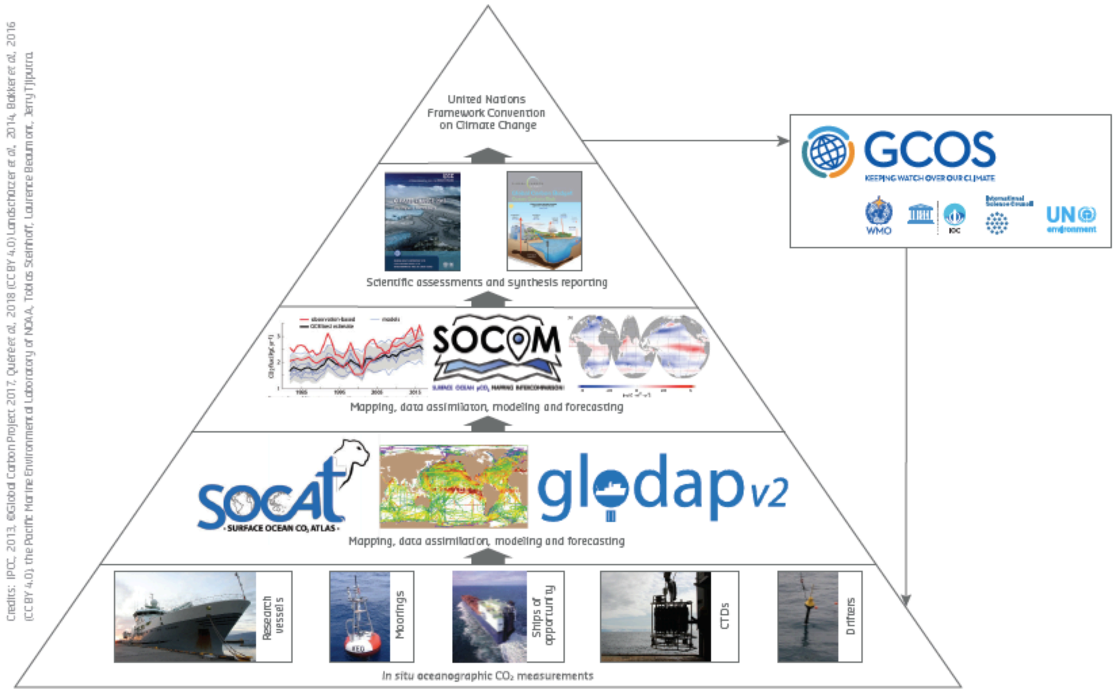

Figure 2. Example of the value chain that connects in situ oceanographic measurements of carbonate chemistry variables to climate negotiations

Nonetheless, the availability and accessibility of these data is not always easy and often the sources are lacking compliance to international standards.

In the European landscape of marine and ocean data management, great progress was made over the last three decades with developing standards, services, and establishing dedicated infrastructures. CMEMS, EMODnet, SDN-NODCs are some of these infrastructures that provide services for the discovery and access to marine in-situ and remote sensing data, ensuring long term stewardship.

Earth Observation technologies are generating an unprecedented high volume of data. Many parameters are measured using different observing platforms and sensors, and there are unique storage and archiving needs for raw data with complex origins. Users of marine data infrastructures require high-quality data and clear source identification.

Traditional data management systems and data-processing software are not specifically designed to handle Big Data. Moreover, traditional systems are not thought to be ready for incorporating new forms of data from datasets that are diverse, large and meant to increase in size (EMB, 2019). The possibility to consume these big data, to collect large volumes of diverse data at a fast pace and to use these data routinely in various marine applications opens up to several interesting scenarios, where setting up the data infrastructure is critical.

A primary challenge is to publish open data in a way that facilitates machine-to-machine data processing. This is driving the infrastructures to improve FAIRness of data and data products in anticipation of the deluge of Big Data and the increasing use of artificial intelligence and cloud-based open science.

Companies like Amazon offer at no cost data hosting, with minimal costs for computer processing. For example, with a relatively small investment of less than €1,000, cloud processing of Big Data – around 400 Terabytes of data – is now possible, using 10s of Terabytes of RAM, thousands of CPUs and using cloud computing and/or cloud-free mosaics.

Nonetheless, potential drawbacks must be considered: while some data are comparable to Big Data (e.g., some physical oceanographic parameters with automated quality control and vessel density tracking data etc.), others are less suitable (e.g., marine biological datasets). In any case, ensuring quality and provenance of data (e.g., though metadata descriptions) and interoperability remains the main challenge.

Moreover, despite the growing capability for near real-time data delivery, there will also remain a need for delayed-mode, validated, harmonised and standardised, integrated datasets, for instance, to validate models.

Research projects have to provide users with the best version of datasets. Hence, it is important to have a private cloud-oriented infrastructure in which data can be hosted and processed before being released to the public.

We therefore provide the following main recommendations:

- Design a big-data cloud-oriented infrastructure
- Adopt standardised procedure for qualifying data
- Implement open data policy
- Adopt open tools for data handling and data publishing
- Implement standards for FAIRness and machine-to-machine interoperability

2.2. Data Handling, Harmonisation and Sharing

Although important progress was made, some problems still exist in the data management framework, particularly related to the integration and harmonisation of data across the multi-scale, multi-platform, multi-sensor ocean observing 'system of systems' (Snowden et al., 2019).

The handling and management of large volumes of heterogeneous data generated from different observation-based infrastructures is a key challenge. IODE published manuals and guides[3,] starting in the 1960s, that dealt with operational procedures for data collection, quality assessment and quality control, standards and reference materials, data formats, etc.

The building block for a distributed ocean observing system in which component systems can exchange and understand information with each other is the standardization of data formats, distribution protocols, and metadata. In the past decade, European partners, in close collaboration with international partners, have been playing an active role in the improvement of environmental data standardisation, accessibility and interoperability through several EU projects (e.g. Copernicus Marine Service - CMEMS, SeaDataNet, AtlantOS, ODIP and EMODnet), enhancing access to observational data at all stages of the data life cycle and fostering the development of integrated services targeted to research, regulatory and operational users.

NAUTILOS is expected to collect, validate and process a huge amount of heterogeneous data that needs dedicated tools and services to facilitate integration and interoperability. As described in the project Data Management Plan (DMP)[4,] NAUTILOS will produce and consume two main data categories: 1) internal project data, i.e., data outputs from the project itself; and 2) externally valuable environmental data, i.e., real-world application data.

[3] https://www.iode.org/index.php?option=com_oe&task=viewDoclistRecord&doclistID=9

[4] D.1.3 Data management Plan

Datasets acquired during the project through sensors and in-situ observation systems will be made freely available to European marine data infrastructures, as well as the international ocean science community and stakeholders. NAUTILOS has to follow and implement some basic data handling, harmonization and sharing recommendations, and as described in D8.2[5], there are three main elements for implementing the interoperability, namely metadata, data formats and data services.

2.2.1. Metadata and controlled vocabularies

'Metadata' refers to vital information about the data collection. Metadata describe a broad range of information that allows observations to be understood and transformed into knowledge. They provide a context for research data, ideally in a machine-readable format, and are used by users' expectations. Metadata require the use of standardised vocabularies to solve ambiguity problems.

Controlled vocabularies consist of lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider scientific community. Examples are the CF conventions for parameters (i.e., parameters standard name), the SeaDataNet/NERC Vocabulary Service (NVS) (i.e., P01-P02-P07). NAUTILOS is going to use and adopt the CF convention and many of the SeaDataNet/NVS vocabularies.

2.2.2. Common data formats

Transport format is one of the key elements for implementing data interoperability and having a common transport format is an open community issue.

The NAUTILOS project involves eight ocean observing platforms with different versions (e.g., the FerryBox platform has four different versions depending on the operator). These observing platforms include: FerryBoxes, autonomous surface vehicles (ASVs), fisheries research vessels, autonomous underwater vehicles (AUVs), Argo profiling floats, deep-sea landers, fixed platforms, and Unmanned Aerial Vehicles (UAVs).

NAUTILOS D.2.2 and D2.3 analysed the technical and interface control requirements. These also include low level data format requirements (e.g., ASCII format) that is different from a higher interoperability layer. While at a lower level the system has to optimise performances (data transfer and space occupancy), at a higher level it is important to complete and complement data with metadata to facilitate their use and re-use.

The European marine data infrastructures are based on the format of the files that are used to distribute OceanSITES data model (with some extensions to provide the user with more information). Typically, the transport format for operational data is NetCDF (CF Convention). The NAUTILOS interoperability layer is recommended to use/offer netCDF (version 4.0) following the data model defined by the EuroGOOS DATAMEQ working group.

This data model includes metadata and information of data quality (flags) and uses a combination of CF convention (current version is CF-1.6) and SeaDataNet controlled vocabularies (e.g. European Directory of Marine Organizations – EDMO, NERC Vocabulary Server for Parameters – P02, ISO8601 for time, WGS84 for date, etc.).

[5] D.8.2 Interoperability requirements definition

The ability to access and search metadata for marine science data is a key requirement for answering fundamental principles of data management, making data FAIR.

Another widely used scientific data file format is HDF[6]. Both NetCDF and HDF provide compact, binary formats optimised for efficient storage and access of large, complex datasets and support features such as internal compression, and support hierarchical structuring of data within files.

2.2.3. Open data publishing catalogues

NAUTILOS is an EU supported research project, hence all the collected data have to be freely available to the community at no cost and limitation. Simply ensuring that data are freely and openly available is not enough to effectively improve data interoperability, though of course this is necessary.

To facilitate the data harmonisation and operate as integrator and data translator for facilitating the data use and interoperability it is important to use common open tools to query for and view data collections and data products. To this end, the GOOS Observation Coordination Group (OCG), coordinating the activities of the global ocean observing networks, is working to improve data interoperability between and within the various observing networks. The OCG is actively promoting the use of ERDDAP as a key tool towards interoperability of global ocean datasets. ERDDAP data server is open-source software written in Java that builds upon the open-source ideals of the OPeNDAP, WCS, SOS and OBIS standards. ERDDAP data server supports several common data file formats (html table, netcdf, csv, txt, mat, json, etc.) and output files are created on-the-fly in any of this format. ERDDAP implements FGDC Web Accessible Folder (WAF) with FGDC-STD-001-1998 and ISO 19115 WAF with ISO 19115-2/19139.

Another tool that is increasing in popularity and use is GeoNetwork. GeoNetwork, based on an open source (GNOS) project, is a free and open source (FOSS) cataloguing application for spatially referenced resources. GeoNetwork provides a web interface to search geospatial data across multiple catalogues. The search provides full-text search as well as faceted search on keywords, resource types, organisations, scale, etc. The catalogue is able to describe geospatial layers, services, maps and also non geographic datasets. Geonetwork implements WxS, OGC, ISO 19115/119/110 standards used for spatial resources and also the Dublin Core format usually used for open data portals.

NAUTILOS has adopted a data interoperability infrastructure that is based on ERDDAP, GeoNetwork and GeoServer (to manage vectorial data and map layers).

We provide the following main recommendations:

- (in situ/ex situ data) File format: NetCDF v.4.0
- Data model: OceanSITES/EuroGOOS DATAMEQ data model
- Metadata:
    - Time: ISO8601 standard "YYYY-MM- DDThh:mm:ssZ" is used; this applies to attributes and to the base date in the ‘units’ attribute for time. UTC must be used, and specified.
    - Latitude and longitude: WGS84
    - Implement Global attributes from - Attribute Convention for Data Discovery (ACDD)
    - Use GEMET-INSPIRE theme
    - Parameters: CF standard names, CF short names and SeaDataNet (SDN) P02/P09

[6] https://www.hdfgroup.org/solutions/hdf5/

  - Institution codes: EDMO (European Directory of Marine Organisations)
  - Country code: ISO 3166
- Data publishing service: ERDDAP + GeoServer
- Data catalogue service: GeoNetwork

2.3. Recommendation for Data Management

The recommendations elaborated for the NAUTILOS data management are here summarised here:

- Design a big-data cloud-oriented infrastructure
- Design a twinned infrastructure to serve both internal (project) users and external users and stakeholders
- Adopt standardised procedure for qualifying data
- Implement open data policy
- Make qualified data available with not additional costs and barriers
- Use open tools for data handling and data publishing
- Implement standards for FAIRness and machine-to-machine interoperability
- Use NetCDF v.4.0 transport format file for sharing in situ observations and measurements
- Use CF 1.6 convention
- Use OceanSITES/EuroGOOS DATAMEQ data model for in situ data
- Complement data with Metadata – as much information as possible
- Implement Global attributes from - Attribute Convention for Data Discovery (ACDD)
- Use GEMET-INSPIRE theme
- Time: ISO8601 standard "YYYY-MM- DDThh:mm:ssZ" is used; this applies to attributes and to the base date in the ‘units’ attribute for time. UTC must be used, and specified.
- Latitude and longitude: WGS84
- Country code: ISO 3166
- Data publishing service: ERDDAP + GeoServer
- Data catalogue service: GeoNetwork

3. NAUTILOS DATA MANAGEMENT

3.1. NAUTILOS DATA VALUE CHAIN

NAUTILOS project will deploy a set of sensors and samplers to measure a series of environmental variables and descriptors essential to understand the state of the ocean, its dynamics and properties, to quantify the forcing of the atmosphere-ocean boundary and to understand the role the oceans play in Earth's climate.

These variables consists of 14 biogeochemical, biological and ecosystem essential ocean variables (EOVs), *i.e.* inorganic carbon, stable carbon isotopes, dissolved oxygen, inorganic macronutrients, suspended particulate, ocean colour, ocean sound, phytoplankton biomass and diversity, zooplankton biomass and diversity, turtles, marine birds, marine mammals abundance and distribution, live coral, seagrass cover, microbial biomass and diversity and invertebrate abundance and distribution, two deep ocean observing system (DOOS) specific EOVs, i.e. litter including micro-plastics, seafloor sponge habitat cover and eight MSFD descriptors.Table 1 provides the list of the NAUTILOS technologies and targeted variables and Figure 3 shows the relationship between EOVs, MSFD and NAUTILOS technologies (see D2.1 for further details).

Table 1. NAUTILOS Sensor Technologies

NAUTILOS Marine Technologies	**Variables targeted**	**Target disciplinary groups**
1. Dissolved Oxygen Sensors (ref. Sub-Task 3.1.1 & Sub-Task 3.1.2)	Dissolved oxygen	Marine biogeochemistry Regulatory environmental monitoring
2. Fluorescence Sensor (ref. Sub-Task 3.1.2)	Chlorophyll-a fluorescence	Marine biology/ecology Regulatory environmental monitoring
3. Ocean surface multi/hyperspectral and laser induced chlorophyll-a fluorescence sensors and cameras (ref. Task 3.2)	Sea surface temperature, Laser induced chlorophyll-a fluorescence, Ocean colour	Marine biology/ecology Regulatory environmental monitoring Ocean colour community
4. Passive broadband acoustic recording sensor (ref. Sub-Task 3.3.1)	Marine noise (anthropogenic and natural sources)	Marine biology/ecology Regulatory environmental monitoring
5. Passive acoustic event recorder (ref. Sub-Task 3.3.2)	Marine mammal sound detection (porpoise & dolphin clicks for abundance estimation)	Marine biology/ecology Regulatory environmental monitoring

6. Active Acoustic Profiling Sensor (ref. Task 3.4)	Suspended particle concentration / distribution (zooplankton, microplastics, organic and inorganic sediment)	Marine biology/ecology Marine pollution
7. Sampler for phytoplankton and other suspended matter (ref. Task 3.5)	Concentrated suspended matter samplers for analyses of phytoplanckon pigments, particulate organic matter, and molecular analysis, microbe biomass and diversity	Marine biology Marine ecological monitoring Climate research Marine pollution
8. Carbonate system/ocean acidification sensors (ref. Task 4.1)	pH, pCO_2,Total Alkalinity, CO_3	Marine biogeochemistry Climate research
9. Silicate Electrochemical Sensor (ref. Task 4.2)	Silicate concentration (Si)	Marine biogeochemistry
10. Submersible Nano- and Microplastics Sampler (ref. Task 4.3)	Concentrated suspended matter samples	Marine ecology Marine pollution
11. Low-cost Microplastic sensors (ref. Task 4.4)	Concentration and characterisation of microplastics	Marine ecology Marine pollution
12. Deep Ocean CTD (ref. Task 4.5)	Conductivity, Temperature, Pressure (Salinity and Density derived)	Physical oceanography
13. Deep ocean low-level radioactivity sensor (ref. Task 4.6)	Radon gas, potassium 40K, radium 226Ra and 228Ra, and other natural isotopes	Environmental monitoring
14. Integration of existing technologie in animal tagging systems (ref. Task 5.5)	Temperature, Salinity, Chlorophyll-a fluorescence, Dissolved oxygen	Physical oceanography Marine biology/ecology
15. Demonstration of novel equipment for key seabed habitat mapping (ref. Task 6.2)	Live corals, hard corals, seafloor sponges	Marine biology
16. Smartphone NIR Scanner (citizen science application, ref. Task 8.4)	Plastics	Marine pollution
17. Visual marine image image annotation (citizen science application, ref. Task 10.4)	Macroplastics, Sponge and cold-water coral cover, major seafloor organism types.	Marine pollution Marine biology/ecology

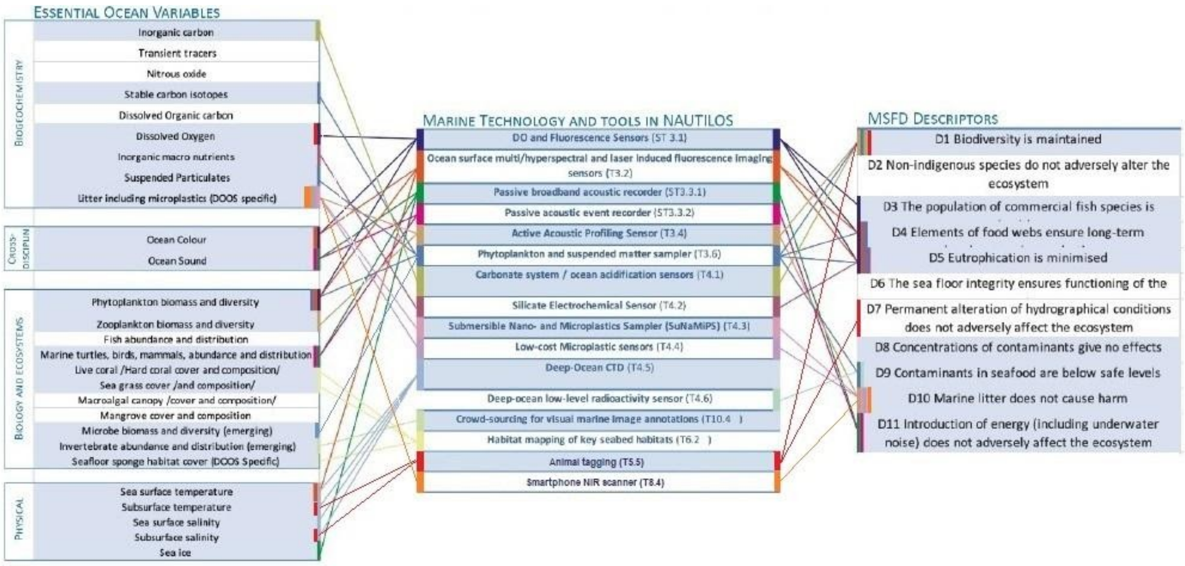

Figure 3 Environmental variables covered by NAUTILOS.

These variables are valuable to various groups and stakeholders ranging from national actors to European Level (MSFD, EMODnet, CMEMS, etc.), and international programs (IODE, GOOS, etc.).

NAUTILOS has already defined its exploitation strategy towards all these stakeholders (see D11.1) and to all of them the possibility to access to more and higher quality data is of fundamental importance (Table 2 see D2.1 for more details) and are ready to ingest these data (Figure 4, see D2.1).

Table 2. Data Gaps for RSC assessment

Eutrophication	**Marine Litter**	**Hazardous substances**
• Limited data availability (both nutrients and chl-a) precludes the assessment in various sub-regions. • Time series are lacking to estimate trends. • More data are needed for the creation of criteria for reference conditions and thresholds/boundary values. • To foster representative data distribution in time and space, monitoring of nutrient concentrations should be increased, especially in or near eutrophication problem areas. • The contribution of the atmospheric nitrogen and phosphorus deposition	Several data issues are reported and these could be improved in following years. • Data are often inconsistent and scarce, making the comparability in sub-regions highly restricted. • More valuable and comparable data could be obtained by standardized approaches (including sampling techniques). • Reference levels and baselines have not been defined. • The lack of long-term data makes the assessment of trends extremely difficult. Longer time series and more monitoring stations are required	The monitoring of Hazardous substances is generally done in the three matrices seawater, sediment and biota. Various gaps are reported, some prevalent in one matrix rather than another. In general: • The observations mainly concern the coastal areas, because the monitoring sites are located here. • There is a lack of data for some subregions, including the most remote areas (for example the Arctic), but not only them. • Existing data are not sufficient to define trends, also to be able to evaluate the effect of measures taken by countries.

cannot be quantified at the basin scale. • The duration of blooms of phytoplankton was difficult to determine owing to restricted sampling. • Assessment of eutrophication status for offshore waters (> 1nm) is still a problem due to the lack of data in these areas.	to detect trends in individual items or types of litter.	Limited spatial coverage, temporal coherence and quality assurance for monitoring activities hinder regional and subregional assessments to some extent

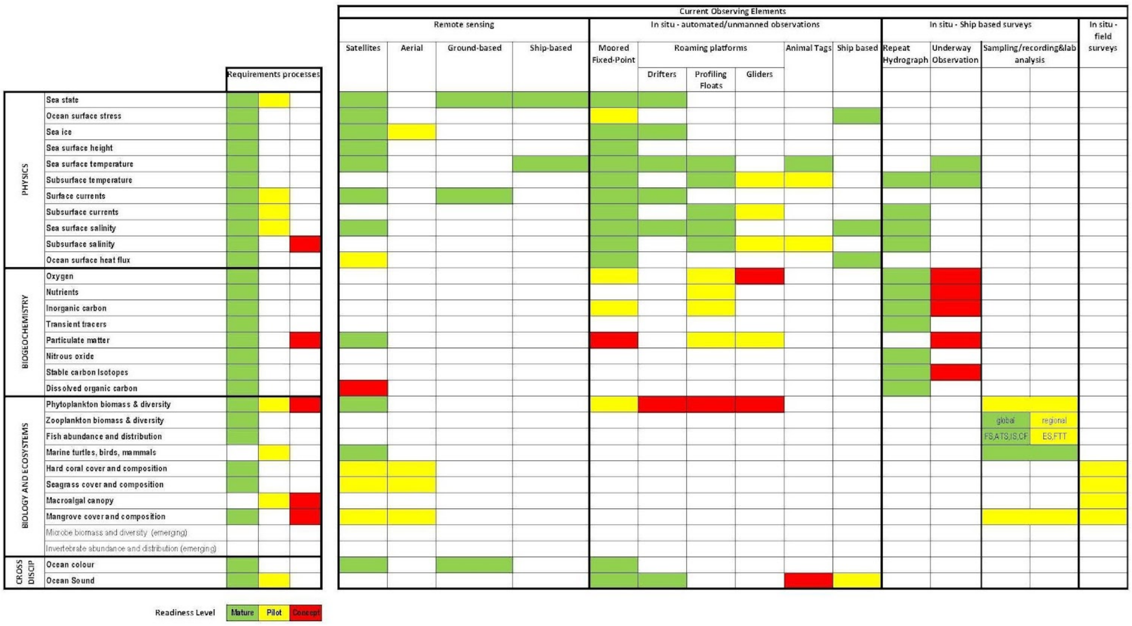

Figure 4. Overview of Essential Ocean Variables (EOVs), their readiness level and the readiness level of their integration in different observing elements/approaches as resulting from the specification sheets (GOOS, 2020).

3.2. NAUTILOS DATA SOURCES AND DATA PLATFORMS

NAUTILOS project is developing new technologies and sensors (see NAUTILOS WP3):

- Fluorometric sensors:
  - dissolved oxygen sensors based on fluorescence quenching
  - dissolved oxygen and chlorophyll-a sensors (profile measurements to 600 m depth)
- Downward-looking multi/hyperspectral and laser induced fluorescence sensors and cameras
- Passive Acoustic Sensors:
  - Noise monitoring
  - Marine fauna abundance estimation
- Active Acoustic Sensors
  - zooplankton biomass abundance quantification
  - microplastic and organic/inorganic sediment distribution
- Phytoplankton and other suspended matter sampler

NAUTILOS is also developing cost-effective sensing technologies in response to the needs for marine chemistry and deep ocean physics measurements, including biogeochemistry EOVs, deep ocean physics EOVs and MSFD descriptors (cf. NAUTILOS WP4). These includes:

- Ocean acidification (impact of the rising CO_2 and declining of ocean pH)
- Silicates
- Nano and microplastics (by means of both picture automatic annotation and fluorescence-based technology)
- Deep ocean CTD and low-level radioactivity sensors

These innovative sensing technologies are going to be plugged in several and different platforms ranging from autonomous vehicles, to opportunity ships and FerryBox, to landers, to fixed mooring buoys and sea mammals (see NAUTILOS WP5).

NAUTILOS D.8.3 presents in detail the approach for managing data for the different platforms from sensors to end-users. In brief, taking the gliders as example (Figure 5), sensors are installed on the platform that is deployed in the ocean. Sensors collect and transmit data in real time and store higher resolution data locally. Some data are shared in real time with various stakeholders via assembly centres (which are interconnected between each other). Once the mission is finished the principal investigator reviews, applies quality checks on collected data, applies common standards and common formats and delivers it to the users-community.

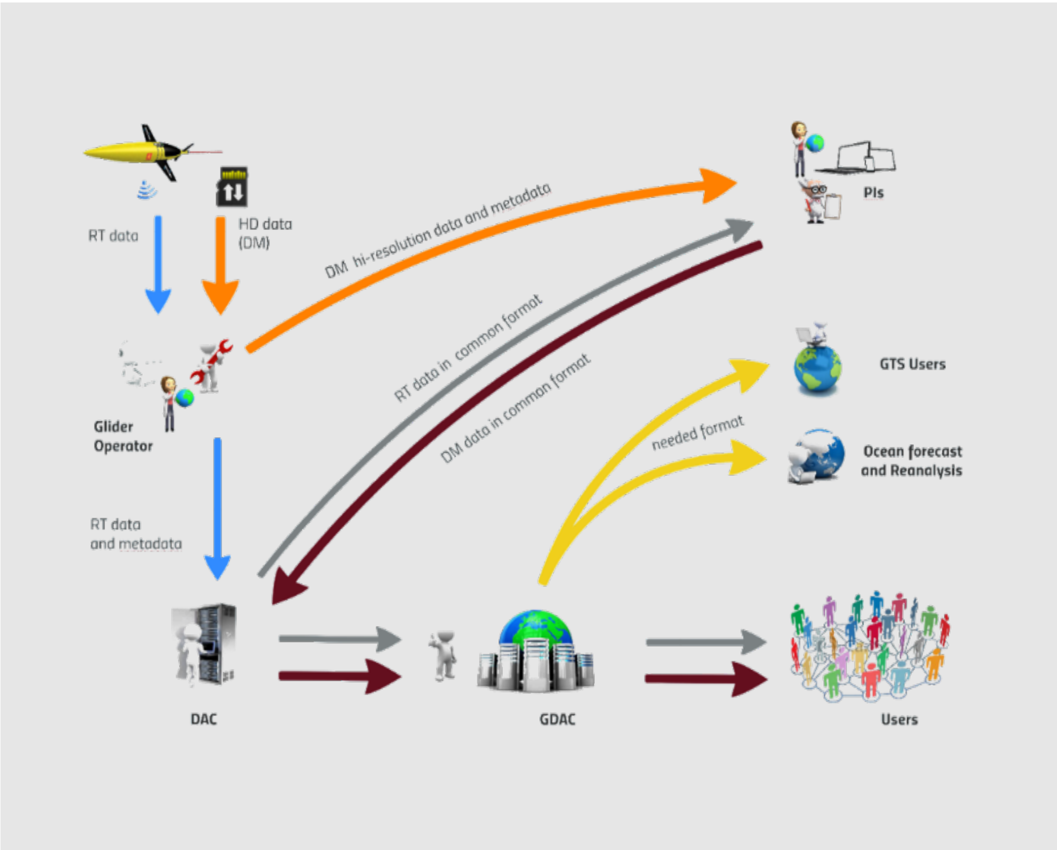

Figure 5. Example of data flow from sensors to users for glider platform (OceanGliders)

3.2. NAUTILOS DATA WORKFLOW AND INFRASTRUCTURE

Following the latest recommendation in data management and to facilitate and enhance the modularity and scalability of the system, NAUTILOS infrastructure is going to have three logical layers (Figure 6), one is designed to manage data and data products, one is designed to organise them to offer the services, and the last one is the application layer i.e. the end-user interface (and features) to access and use the developed and provided services.

Each layer operates by means of a series of microservices in charge of specific actions, e.g. the data layer operates a specific microservice for each source that harvests and completes the metadata, collects data, uses a common transport format, etc.

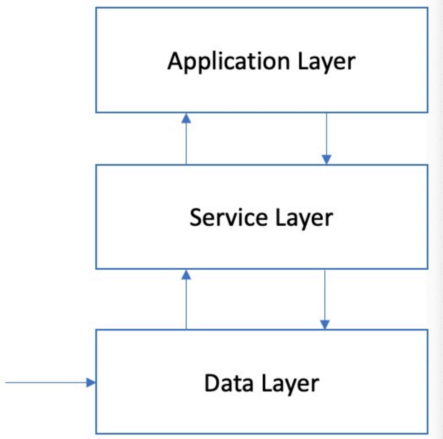

Figure 6. NAUTILOS infrastructure logical layers

The data layer is in charge for ingesting datasets and is able to make available data in different formats, like csv, txt, netcdf, etc.

In data management, it is important to verify the quality of the sampled data. This could be carried out automatically or manually, depending on data and instrument type. NAUTILOS data consists of a combination of measurements of different parameters and images gathered by innovative sensors. Therefore, most of the NAUTILOS quality control has to be performed manually (Table 3).

NAUTILOS applies good practices for FAIRness and follows the general guidelines for data management[7]. Data may become available to all users immediately and, to offer the best quality, data are fully released once quality control is applied.

NAUTILOS data infrastructure manages two levels of data disclosure rights: a private level that is open to partners only for internal data exchange and application of needed quality checks, and a public level that receives data from the private level and makes data open to all the NAUTILOS stakeholders.

Table 3. Sensors and standards reference

Sensor(s)	Standard reference
Deep-ocean CTD	http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf

[7]https://repository.oceanbestpractices.org/bitstream/handle/11329/275/IOC%20Manual_Guides_73.pdf?sequence=1&isAllowed=y

Silicate electrochemical sensors	EMODnet Chemistry eutrophication
Phytoplankton and suspended matter sampler	https://obis.org/manual/
Dissolved oxygen and fluorescence sensor	http://www.argodatamgt.org/Documentation
Surface multi-hyperspectral and laser-induced fluorescence imaging sensors	To be defined
Ocean acidification sensors	EMODnet Chemistry Ocean Acidification
Submersible nano- and microplastics sampler Low-cost microplastic sensors	https://www.emodnet-chemistry.eu/repository/Proposal-EMODnet-TG-ML-Micro-Litter-Data-Gathering-03062020.pdf
Passive broadband acoustic recorder Passive acoustic event recorder	https://publications.jrc.ec.europa.eu/repository/handle/JRC88045 https://obis.org/manual/
Active acoustic profiling sensor	To be defined
Crowd-sourcing for visual marine image annotations	https://obis.org/manual/ http://obis.org/manual/darwincore/
Habitat mapping of key seabed habitats	https://obis.org/manual/ https://www.emodnet-seabedhabitats.eu/contribute-data/data-exchange-format/#h8ee264581b0449caabb16bfdf66420f7

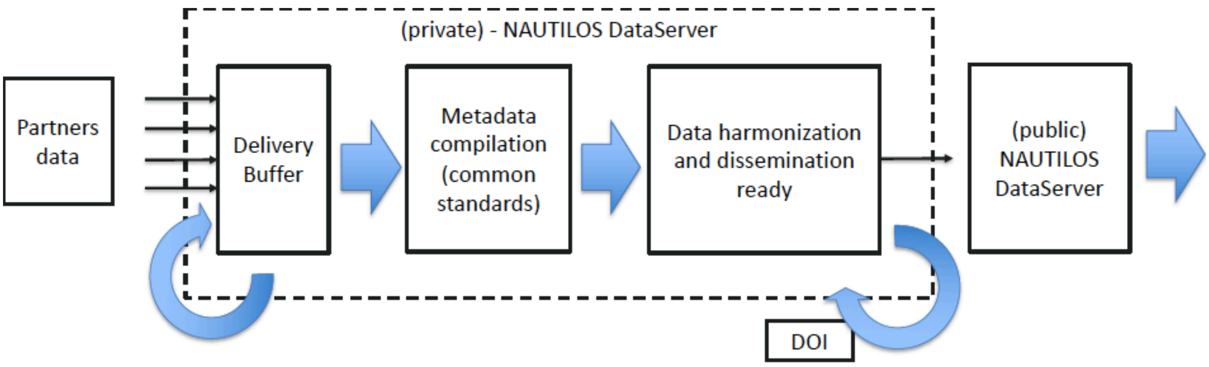

Figure 7. NAUTILOS designed data layer workflow

Figure 7 shows a simplified schema of the data workflow: NAUTILOS infrastructure offers the partner a framework to exchange data, cross validate new sensors and methodologies. Data and datasets are then harmonized according to common standards and made ready for dissemination (service layer).

At this level, as recommended by the IOC data management plan manual, (some) datasets can be stored individually and be given a DOI (a unique, persistent digital identifier). Notably a DOI is also important when datasets are used from large data bases that are subjected to changes, e.g., annual reprocessing, versioning, etc.

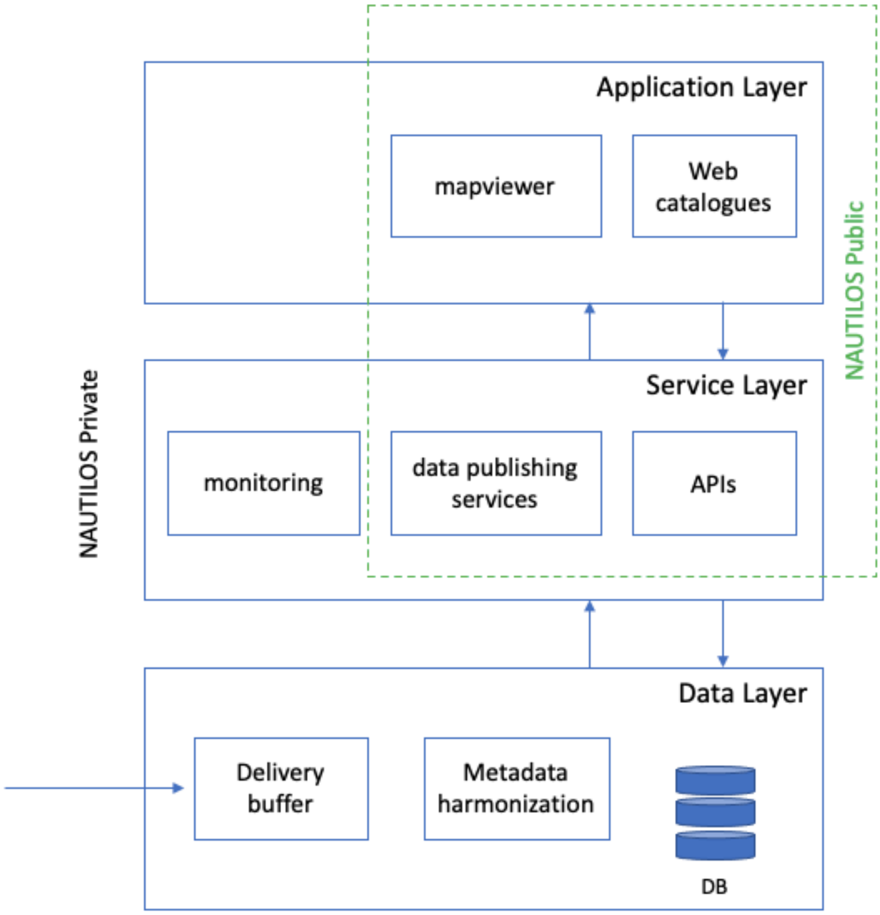

Figure 8. NAUTILOS infrastructure layers and modules

A single dataset source may be integrated in more products and the service layer organizes data within the NAUTILOS datasets and data products publication services and catalogues (ERDDAP, GeoServer, GeoNetwork). The application layer implements the NAUTILOS interfaces (map portal) and other machine-to-machine services (Figure 8).

3.3. DEPLOYED ARCHITECTURE

NAUTILOS VMs are deployed at the Ligurian District for Marine Technologies (DLTM) High Performance Computing (HPC) Lab that is equipped with 504 Cores and 1TB of RAM for Parallel Computing and 240 cores and 1.5 TB of RAM in Cloud, with the aim to facilitating access to state-of-the-art computing infrastructures for DLTM members8 and Ligurian companies. The current

[8] Some NAUTILOS partners are DLTM members

NAUTILOS backend consists of 2 VMs: 2.10GHz - 4 Core, 32GB, running CentOS and 500GB of allocated disk size. The VMs are running Docker containers.

Containers are similar to VMs, but they have relaxed isolation properties to share the Operating System (OS) amongst the applications. Similar to a VM, a container has its own file system, share of CPU, memory, process space, and more. As they are decoupled from the underlying infrastructure, they are portable across clouds and OS distributions.

The most popular solutions are Kubernetes and Docker that have become the de- facto standard and dominant tool with which applications are containerized, managed, scaled and released. As anticipated, NAUTILOS adopted Docker that offers packages with service tools like selected data publishing services.

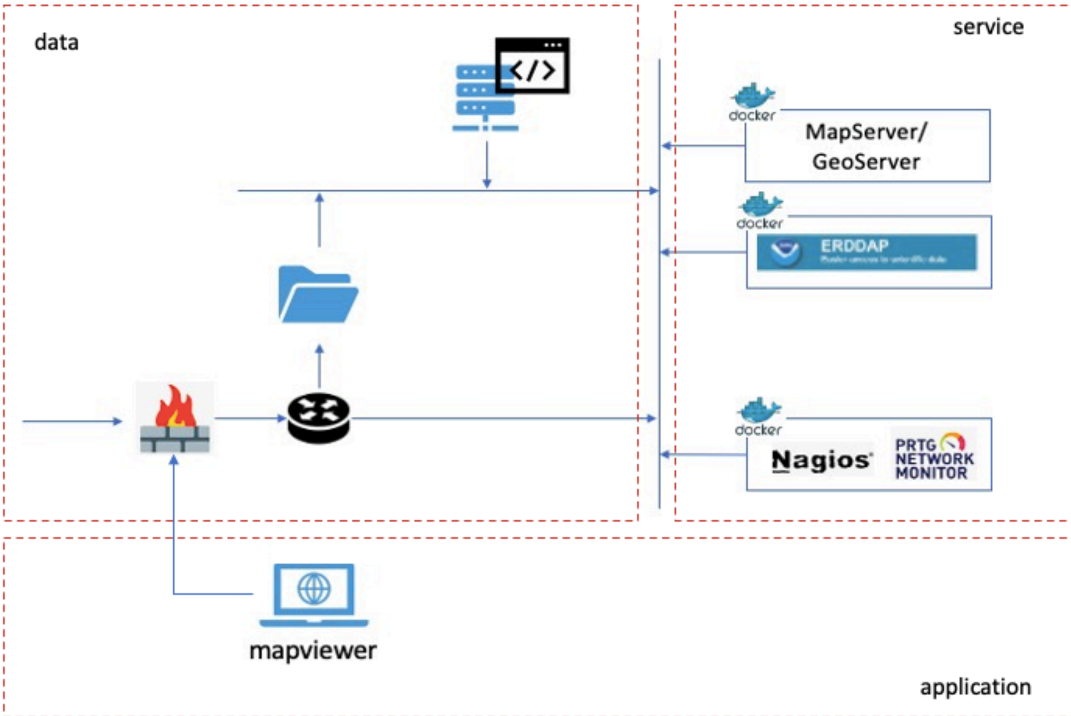

Figure 9. Deployed architecture

As shown in Figure 9, the management of the access to the private area is implemented by using a selective endpoint access from a given (white) list of IP address or host and domain name (fully qualified domain name, FQDN). Connections and access from edge devices out of the list will be rejected.

The idea to design a delivery buffer is mutated from the CMEMS Dissemination Unit (DU)[9] Delivery buffer. Similarly, the NAUTILOS Delivery buffer provides the partners with virtual folders in which new data can be dropped. As described in the NAUTILOS DMP, a preliminary approach suggests the

[9] The CMEMS DU is in charge of the dissemination of the products elaborated by Production Centres (PC; MFCs and TACs), and provide the users with available products by means of several dissemination interfaces. The CMEMS DU is also in charge for monitoring overall system performances (timeliness, availability, bandwidth load, etc.).

partners to use an internal naming convention for new datasets. This facilitates the successive ingestion step that pushes new data into the NAUTILOS database and the data publication tools.

As anticipated, NAUTILOS data publishing identified tools are ERDDAP and GeoServer, which are tools for both the NAUTILOS service layer (they offer application program interfaces and machine-to-machine services for internal use) and the NAUTILOS application layer (they offer cataloguing features and machine-to-machine tools for external users).

The application layer also includes a GeoNetwork instance and the three catalogues are available at the following endpoints:

- erddap.nautilos-h2020.eu
- geoserver.nautilos-h2020.eu
- geonetwork.nautilos-h2020.eu

These interfaces are using the minimum set of metadata as defined in Table 4 (for more details see NAUTILOS deliverable D.8.2).

Table 4. Metadata Vocabularies

Metadata field	**Reference**	**Vocabulary governance**
platform_type	http://vocab.nerc.ac.uk/collection/L06/current/	BODC
platform_type_bigram	CMEMS INSTAC	EuroGOOS DATAMEQ
contributors_role		NAUTILOS
naming_authority	https://edmo.seadatanet.org/	SeaDataNet
institution	https://edmo.seadatanet.org/	SeaDataNet
country	ISO3166	
ICES_code	https://ocean.ices.dk/codes/ShipCodes.aspx	ICES
sensor_model	http://vocab.nerc.ac.uk/collection/L22/current/	BODC
data_mode	RT/DM/REP	EuroGOOS DATAMEQ
variable names	http://vocab.nerc.ac.uk/collection/P02/current/ http://vocab.nerc.ac.uk/collection/P01/current/ http://vocab.nerc.ac.uk/collection/P07/current/	BODC
time	ISO8601	ISO
Date	WGS84	ISO
taxon	LSID	WoRMS

These endpoints represent a very easy tool to serve and interoperate with NAUTILOS stakeholders. The following paragraphs report some examples on how to query for data from these interfaces. At the time of writing this document, NAUTILOS is not exposing new data yet, hence examples are provided by other infrastructures (e.g. erddap.emodnet-physics.eu or coastwatch.pfeg.noaa.gov/erddap/). This will be available for NAUTILOS as soon as preliminary data are generated.

3.3.1. ERDDAP

Tabledap requests a data subset, a graph, or a map from a tabular dataset (for example, buoy data), via a specially formed URL.

Example:

https://erddap.nautilos-h2020.eu/tabledap/*datasetID.fileType*{?*query*}

The query is often a comma-separated list of desired variable names, followed by a collection of constraints (e.g., *variable*<*value*), each preceded by '&' (which is interpreted as "AND").

Details: Requests must not have any internal spaces, Requests are case sensitive, {} is notation to denote an optional part of the request, **datasetID** identifies the name that ERDDAP assigned to the source website and dataset (for example, pmelTaoDySst), **fileType** specifies the type of table data file that you want to download (for example, .htmlTable).

JSON responses from ERDDAP (metadata, gridded data, and tabular/in-situ data) use the same basic format: a database-like table. Since data from EDDTable datasets is already a table (with a column for each requested variable), ERDDAP can easily store the data in a .json file.

Example


```
{
  "table": {
    "columnNames": ["longitude", "latitude", "time", "bottle_posn", "temperature1"],
    "columnTypes": ["float", "float", "String", "byte", "float"],
    "columnUnits": ["degrees_east", "degrees_north", "UTC", null, "degree_C"],
    "rows": [
      [-124.82, 42.95, "2002-08-17T00:49:00Z", 1, 8.086],
      [-124.82, 42.95, "2002-08-17T00:49:00Z", 2, 8.585],
      [-124.82, 42.95, "2002-08-17T00:49:00Z", 3, 8.776],
      ...
      [-124.1, 44.65, "2002-08-19T20:18:00Z", 3, null]
    ]
  }
}
```

All.json responses from ERDDAP have a table object (with name=value pairs), a columnNames, columnTypes, and columnUnits array, with a value for each column, a rows array of arrays with the rows and columns of data, null's for missing values.

MATLAB, users can use tabledap's .mat file type to download data from within MATLAB. Example:

```
load(urlwrite('https://erddap.nautilos-h2020.eu/tabledap/pmelTaoDySst.mat?time,T_25&station="0n0e"&time>=2015-05-23T12:00:00Z&time<=2015-05-31T12:00:00Z&.draw=lines', 'test.mat'));
```

ERDDAPY is a Python library that "takes advantage of ERDDAP's RESTful web services and creates the ERDDAP URL for any request like searching for datasets, acquiring metadata, downloading data, etc." To install erddapy:

```
!pip install erddapy

!pip install erddap-python
```

Example:

```
from erddapClient import ERDDAP_Tabledap

Dataset = ‘ xxx’

remote = ERDDAP_Tabledap('http://erddap.nautilos-h2020.eu /’, Dataset’)

Set_time = …

data = (remote.setResultVariables([‘XXX’, ‘YYY’]).addConstraint(set_time).getDataFrame())
```

Where XXX, YYYY are the varialble names.

3.3.2. GEOSERVER

Geoserver exposes OGC compliant interoperability features e.g. to consume a WMS layer the call would be:

```
wmsOptions={
  layers: "nautilos:LAYERNAME",
  style: "line",
  format:"image/png",
  transparent: "true"
};
var wmsLayer = L.tileLayer.wms("http://geoserver.nautilos-h2020.eu/geoserver/nautilos/LAYERNAME/wms?", wmsOptions).addTo(map);
```

3.3.3. GEONETWORK

One of the most useful GeoNetwork features is the web catalogue service (WCS) that exposes details on how to query data in the catalogue. The user can collect the feature by the getCapabilities:

https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetCapabilities

and use the xml to link and use the NAUTILOS catalogue[10]:


```xml
<csw:Capabilities xmlns:csw="http://www.opengis.net/cat/csw/2.0.2" xmlns:gml="http://www.opengis.net/gml" xmlns:ows="http://www.opengis.net/ows" xmlns:ogc="http://www.opengis.net/ogc" xmlns:xlink="http://www.w3.org/1999/xlink" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" version="2.0.2" xsi:schemaLocation="http://www.opengis.net/cat/csw/2.0.2 http://schemas.opengis.net/csw/2.0.2/CSW-discovery.xsd">
<ows:ServiceIdentification>
<ows:Title>NAUTILOS catalogue</ows:Title>
<ows:Abstract/>
<ows:Keywords>
<!--  Keywords are automatically added by GeoNetwork
      according to catalogue content.  -->
<ows:Keyword>Oceanographic geographical features</ows:Keyword>
<ows:Keyword>latitude</ows:Keyword>
<ows:Keyword>longitude</ows:Keyword>
<ows:Keyword>time</ows:Keyword>
<ows:Keyword>data</ows:Keyword>
<ows:Keyword>sea</ows:Keyword>
<ows:Keyword>ocean</ows:Keyword>
<ows:Keyword>science</ows:Keyword>
<ows:Keyword>seawater</ows:Keyword>
<ows:Keyword>oceans</ows:Keyword>
<ows:Type>theme</ows:Type>
</ows:Keywords>
<ows:ServiceType>CSW</ows:ServiceType>
<ows:ServiceTypeVersion>2.0.2</ows:ServiceTypeVersion>
<ows:Fees/>
<ows:AccessConstraints/>
</ows:ServiceIdentification>
<ows:ServiceProvider>
<ows:ProviderName/>
<ows:ProviderSite xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork"/>
<ows:ServiceContact>
<ows:IndividualName/>
<ows:PositionName/>
```


[10] At the time of writing the catalogue was not available. The example is based on the expected system answer.

```xml
<ows:ContactInfo>
<ows:Phone>
<ows:Voice/>
<ows:Facsimile/>
</ows:Phone>
<ows:Address>
<ows:DeliveryPoint/>
<ows:City/>
<ows:AdministrativeArea/>
<ows:PostalCode/>
<ows:Country/>
<ows:ElectronicMailAddress/>
</ows:Address>
<ows:HoursOfService/>
<ows:ContactInstructions/>
</ows:ContactInfo>
<ows:Role>pointOfContact</ows:Role>
</ows:ServiceContact>
</ows:ServiceProvider>
<ows:OperationsMetadata>
<ows:Operation name="GetCapabilities">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
</ows:HTTP>
</ows:DCP>
<ows:Parameter name="sections">
<ows:Value>ServiceIdentification</ows:Value>
<ows:Value>ServiceProvider</ows:Value>
<ows:Value>OperationsMetadata</ows:Value>
<ows:Value>Filter_Capabilities</ows:Value>
</ows:Parameter>
<ows:Constraint name="PostEncoding">
<ows:Value>XML</ows:Value>
<ows:Value>SOAP</ows:Value>
</ows:Constraint>
</ows:Operation>
```

```
<ows:Operation name="DescribeRecord">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
</ows:HTTP>
</ows:DCP>
<ows:Parameter name="outputFormat">
<ows:Value>application/xml</ows:Value>
</ows:Parameter>
<ows:Parameter name="schemaLanguage">
<ows:Value>http://www.w3.org/TR/xmlschema-1/</ows:Value>
</ows:Parameter>
<ows:Parameter xmlns:gfc="http://www.isotc211.org/2005/gfc" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:mdb="http://standards.iso.org/iso/19115/-3/mdb/2.0" name="outputSchema">
<ows:Value>http://www.opengis.net/cat/csw/2.0.2</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gfc</ows:Value>
<ows:Value>http://www.w3.org/ns/dcat#</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gmd</ows:Value>
<ows:Value>http://standards.iso.org/iso/19115/-3/mdb/2.0</ows:Value>
</ows:Parameter>
<ows:Constraint name="PostEncoding">
<ows:Value>XML</ows:Value>
<ows:Value>SOAP</ows:Value>
</ows:Constraint>
</ows:Operation>
<ows:Operation name="GetDomain">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
</ows:HTTP>
</ows:DCP>
</ows:Operation>
<ows:Operation name="GetRecords">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
```

```
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
</ows:HTTP>
</ows:DCP>
<!--  FIXME : Gets it from enum or conf  -->
<ows:Parameter name="resultType">
<ows:Value>hits</ows:Value>
<ows:Value>results</ows:Value>
<ows:Value>validate</ows:Value>
</ows:Parameter>
<ows:Parameter name="outputFormat">
<ows:Value>application/xml</ows:Value>
</ows:Parameter>
<ows:Parameter xmlns:gfc="http://www.isotc211.org/2005/gfc" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:mdb="http://standards.iso.org/iso/19115/-3/mdb/2.0" name="outputSchema">
<ows:Value>http://www.opengis.net/cat/csw/2.0.2</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gfc</ows:Value>
<ows:Value>http://www.w3.org/ns/dcat#</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gmd</ows:Value>
<ows:Value>http://standards.iso.org/iso/19115/-3/mdb/2.0</ows:Value>
</ows:Parameter>
<ows:Parameter xmlns:gfc="http://www.isotc211.org/2005/gfc" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:mdb="http://standards.iso.org/iso/19115/-3/mdb/2.0" name="typeNames">
<ows:Value>csw:Record</ows:Value>
<ows:Value>gfc:FC_FeatureCatalogue</ows:Value>
<ows:Value>dcat</ows:Value>
<ows:Value>gmd:MD_Metadata</ows:Value>
<ows:Value>mdb:MD_Metadata</ows:Value>
</ows:Parameter>
<ows:Parameter name="CONSTRAINTLANGUAGE">
<ows:Value>FILTER</ows:Value>
<ows:Value>CQL_TEXT</ows:Value>
</ows:Parameter>
<ows:Constraint name="PostEncoding">
<ows:Value>XML</ows:Value>
<ows:Value>SOAP</ows:Value>
</ows:Constraint>
<ows:Constraint name="SupportedISOQueryables">
```

<ows:Value>CreationDate</ows:Value>
<ows:Value>GeographicDescriptionCode</ows:Value>
<ows:Value>OperatesOn</ows:Value>
<ows:Value>Modified</ows:Value>
<ows:Value>DistanceUOM</ows:Value>
<ows:Value>Operation</ows:Value>
<ows:Value>ResourceIdentifier</ows:Value>
<ows:Value>Format</ows:Value>
<ows:Value>Identifier</ows:Value>
<ows:Value>Language</ows:Value>
<ows:Value>ServiceType</ows:Value>
<ows:Value>OrganisationName</ows:Value>
<ows:Value>KeywordType</ows:Value>
<ows:Value>AnyText</ows:Value>
<ows:Value>PublicationDate</ows:Value>
<ows:Value>AlternateTitle</ows:Value>
<ows:Value>Abstract</ows:Value>
<ows:Value>HasSecurityConstraints</ows:Value>
<ows:Value>Title</ows:Value>
<ows:Value>CouplingType</ows:Value>
<ows:Value>TopicCategory</ows:Value>
<ows:Value>ParentIdentifier</ows:Value>
<ows:Value>Subject</ows:Value>
<ows:Value>ResourceLanguage</ows:Value>
<ows:Value>TempExtent_end</ows:Value>
<ows:Value>ServiceTypeVersion</ows:Value>
<ows:Value>Type</ows:Value>
<ows:Value>RevisionDate</ows:Value>
<ows:Value>OperatesOnName</ows:Value>
<ows:Value>Denominator</ows:Value>
<ows:Value>DistanceValue</ows:Value>
<ows:Value>TempExtent_begin</ows:Value>
<ows:Value>OperatesOnIdentifier</ows:Value>
</ows:Constraint>
<ows:Constraint name="AdditionalQueryables">
<ows:Value>SpecificationDate</ows:Value>
<ows:Value>AccessConstraints</ows:Value>
<ows:Value>ResponsiblePartyRole</ows:Value>

```
<ows:Value>Degree</ows:Value>
<ows:Value>Lineage</ows:Value>
<ows:Value>OnlineResourceMimeType</ows:Value>
<ows:Value>ConditionApplyingToAccessAndUse</ows:Value>
<ows:Value>Date</ows:Value>
<ows:Value>MetadataPointOfContact</ows:Value>
<ows:Value>OnlineResourceType</ows:Value>
<ows:Value>Relation</ows:Value>
<ows:Value>SpecificationDateType</ows:Value>
<ows:Value>Classification</ows:Value>
<ows:Value>OtherConstraints</ows:Value>
<ows:Value>SpecificationTitle</ows:Value>
</ows:Constraint>
</ows:Operation>
<ows:Operation name="GetRecordById">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw"/>
</ows:HTTP>
</ows:DCP>
<ows:Parameter xmlns:gfc="http://www.isotc211.org/2005/gfc" xmlns:dcat="http://www.w3.org/ns/dcat#" xmlns:gmd="http://www.isotc211.org/2005/gmd" xmlns:mdb="http://standards.iso.org/iso/19115/-3/mdb/2.0" name="outputSchema">
<ows:Value>http://www.opengis.net/cat/csw/2.0.2</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gfc</ows:Value>
<ows:Value>http://www.w3.org/ns/dcat#</ows:Value>
<ows:Value>http://www.isotc211.org/2005/gmd</ows:Value>
<ows:Value>http://standards.iso.org/iso/19115/-3/mdb/2.0</ows:Value>
</ows:Parameter>
<ows:Parameter name="outputFormat">
<ows:Value>application/xml</ows:Value>
</ows:Parameter>
<ows:Parameter name="resultType">
<ows:Value>hits</ows:Value>
<ows:Value>results</ows:Value>
<ows:Value>validate</ows:Value>
</ows:Parameter>
<ows:Parameter name="ElementSetName">
```

```xml
<ows:Value>brief</ows:Value>
<ows:Value>summary</ows:Value>
<ows:Value>full</ows:Value>
</ows:Parameter>
<ows:Constraint name="PostEncoding">
<ows:Value>XML</ows:Value>
<ows:Value>SOAP</ows:Value>
</ows:Constraint>
</ows:Operation>
<ows:Operation name="Transaction">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw-publication"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw-publication"/>
</ows:HTTP>
</ows:DCP>
</ows:Operation>
<ows:Operation name="Harvest">
<ows:DCP>
<ows:HTTP>
<ows:Get xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw-publication"/>
<ows:Post xlink:href="https://geonetwork.nautilos-h2020.eu/geonetwork/srv/eng/csw-publication"/>
</ows:HTTP>
</ows:DCP>
<ows:Parameter name="ResourceType">
<ows:Value>http://www.isotc211.org/schemas/2005/gmd/</ows:Value>
</ows:Parameter>
</ows:Operation>
<ows:Parameter name="service">
<ows:Value>http://www.opengis.net/cat/csw/2.0.2</ows:Value>
</ows:Parameter>
<ows:Parameter name="version">
<ows:Value>2.0.2</ows:Value>
</ows:Parameter>
<ows:Constraint name="IsoProfiles">
<ows:Value>http://www.isotc211.org/2005/gmd</ows:Value>
</ows:Constraint>
```

```xml
<ows:Constraint name="PostEncoding">
<ows:Value>SOAP</ows:Value>
</ows:Constraint>
</ows:OperationsMetadata>
<ogc:Filter_Capabilities>
<ogc:Spatial_Capabilities>
<ogc:GeometryOperands>
<ogc:GeometryOperand>gml:Envelope</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:Point</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:LineString</ogc:GeometryOperand>
<ogc:GeometryOperand>gml:Polygon</ogc:GeometryOperand>
</ogc:GeometryOperands>
<ogc:SpatialOperators>
<ogc:SpatialOperator name="BBOX"/>
<ogc:SpatialOperator name="Equals"/>
<ogc:SpatialOperator name="Overlaps"/>
<ogc:SpatialOperator name="Disjoint"/>
<ogc:SpatialOperator name="Intersects"/>
<ogc:SpatialOperator name="Touches"/>
<ogc:SpatialOperator name="Crosses"/>
<ogc:SpatialOperator name="Within"/>
<ogc:SpatialOperator name="Contains"/>
<!--
    <ogc:SpatialOperator name="Beyond"/>
    <ogc:SpatialOperator name="DWithin"/>
     The 'SpatialOperator' element can have a GeometryOperands child  -->
</ogc:SpatialOperators>
</ogc:Spatial_Capabilities>
<ogc:Scalar_Capabilities>
<ogc:LogicalOperators/>
<ogc:ComparisonOperators>
<ogc:ComparisonOperator>EqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>Like</ogc:ComparisonOperator>
<ogc:ComparisonOperator>LessThan</ogc:ComparisonOperator>
<ogc:ComparisonOperator>GreaterThan</ogc:ComparisonOperator>
<!--  LessThanOrEqualTo is in OGC Filter Spec, LessThanEqualTo is in OGC CSW schema  -->
<ogc:ComparisonOperator>LessThanEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>LessThanOrEqualTo</ogc:ComparisonOperator>
```

```
<!--  GreaterThanOrEqualTo is in OGC Filter Spec, GreaterThanEqualTo is in OGC CSW schema  -->
<ogc:ComparisonOperator>GreaterThanEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>GreaterThanOrEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>NotEqualTo</ogc:ComparisonOperator>
<ogc:ComparisonOperator>Between</ogc:ComparisonOperator>
<ogc:ComparisonOperator>NullCheck</ogc:ComparisonOperator>
<!--  FIXME : Check NullCheck operation is available  -->
</ogc:ComparisonOperators>
</ogc:Scalar_Capabilities>
<ogc:Id_Capabilities>
<ogc:EID/>
<ogc:FID/>
</ogc:Id_Capabilities>
</ogc:Filter_Capabilities>
</csw:Capabilities>
```

4. CONCLUSIONS AND PERSPECTIVES

NAUTILOS data management infrastructure sets up the workflow for delivering information and products to both internal and external users. The data workflow is based on a chain of individual operational components, each contributing either to data production or user interaction functions.

The infrastructure implements open and free of charge data access, it implements international standards, follows European regulations such as INSPIRE, it is accessible under the NAUTILOS web domain (www.nautilos-h2020.eu) and offers up to date means of data and data products dissemination and interaction tools.

Nevertheless, the developed system cannot be considered finally closed, and for the entire duration of the project, the following evolution principles have to be considered and applied:

- **The service shall be state-of-the-art in its scope, and it shall be user-driven.**

This means that all service capabilities are defined, allocated and implemented with sound and state-of-the-art technical and scientific justifications and address NAUTILOS internal and external user needs. If new needs pop up, it is important to try and match (whenever possible) these new needs.

Providing international level state-of-the-art tools means a constant monitoring and application of most appropriate methodologies at a given time to meet requirements for service quality in data processing, validation, etc.

- **Common standards are adopted and applied throughout the different elements of the system.**

This means that standards are identified, adopted and applied. If, during the lifetime of the project, rising ocean best practices (OBP) become standards, the project has to try and include (as much as possible) these new OBP.

This also means that the system evolves constantly in response to service requirements, scientific and technical requirements, marine environment products requirements, stakeholders' requirements.

- **Quality of the methodology of workflow has to follow and apply, well-established engineering methodologies derived from industry (es. ISO9001:2015, ISO15288, etc.) drive the definition of an adapted approach.**

This means that data management also adopts good practices and methodologies to ensure system monitoring, fixing and reporting.

Within this framework, NAUTILOS poses itself amongst the up-to-date data infrastructure projects, adopts the latest interoperability best practices and enables data-products FAIRness to maximise the NAUTILOS data value chain from production to its consume for added value ocean products.

Appendix 1: References and Related Documents

ID	Reference or Related Document	Source or Link/Location
1	NAUTILOS Grant Agreement No. 101000825	*NAUTILOS ownCloud < Confidential document >*
2	NAUTILOS D.1.3 - DMP	*10.5281/zenodo.7163625*
3	NAUTILOS D.2.1	*10.5281/zenodo.7163906*
4	NAUTILOS D.2.2	*NAUTILOS ownCloud < Confidential document >*
5	NAUTILOS D.2.3	*NAUTILOS ownCloud < Confidential document >*
6	NAUTILOS D.8.2	*10.5281/zenodo.7211792*
7	NAUTILOS D.11.1	*10.5281/zenodo.7211802*
8	European Marine Board Annual Report 2019 - EMB, 2019	*https://www.marineboard.eu/publications/emb-annual-report-2019*
9	Snowden et al., 2019	*https://doi.org/10.3389/fmars.2019.00442*