www.nautilos-h2020.eu

Organisation CNR
Department ISTI

Data Management Plan

2nd periodic report update

Date: 17.10.2023
Doc. Version: V1.3
10.5281/zenodo.10664813

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101009825 (NAUTILOS). This output reflects only the author's

Document Control Information

Settings	Value
Deliverable Title	Data Management Plan – 2nd periodic report update
Work Package Title	Project Management
Deliverable number	D1.11
Description	In this deliverable an update of the DMP at M36 will be reported.
Lead Beneficiary	CNR
Lead Authors	Antonio Novellino
Contributors	Gabriele Pieri, Francesco Misurale
Submitted by	Gabriele Pieri
Doc. Version (Revision number)	V.1.3
Sensitivity (Security):	Public
Date:	17/10/2023

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
Sandra Sá	Reviewer	*Approved*	17/10/2023

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
V1.1	05/09/23	Antonio Novellino	Updated doc
V1.2	13/10/23	Gabriele Pieri	Final version ready for review
V1.3	17/10/23	Gabriele Pieri	Finalised version after review

Configuration Management: Document Location

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

Nature of the deliverable		
R	Report	
DEC	Websites, patents, filing, etc.	
DEM	Demonstrator	
O	Other	
ORDP	Open Research Data Pilot	✓

Dissemination level		
PU	Public	✓
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: https://www.nautilos-h2020.eu/.

COPYRIGHT

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The timely, free and unrestricted exchange of oceanographic observational data is essential for the efficient acquisition, integration and use of ocean observations gathered by the several projects carried out all over the world for a variety of purposes, such as weather forecasts and climate projections, the preservation of wildlife, marine and coastal environmental management etc.

An open and free data policy is highly promoted by the European Commission and its member states for a wide range of environmental data services targeted to a wide range of user communities. Interoperability of data systems has become a priority with the development of FAIR principles[1], *i.e.,* a set of guiding principles to make data **Findable, Accessible, Interoperable, and Re-usable**.

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.

In line with these recommendations and agreements, NAUTILOS aims to make accessible and freely available on the internet all the marine data gathered within the project. The NAUTILOS data management policy is committed to make available NAUTILOS data interoperable with other European and Global ocean data initiatives, more specifically it is targeting key EU marine data integration programs such as EMODnet with its thematic lots, CMEMS (in particular the INSTAC), SeaDataNet network of NODCs, etc. NAUTILOS is developing and deploying state of art data management backend infrastructure and tools and these support machine to machine interoperability features to be used by NAUTILOS stakeholder to harvest and use relevant data and products. Furthermore, NAUTILOS data management policy clarifies the roles on the ownership and custodianship of the data, as well as the recommendations on data citation.

Data sharing and dissemination principles:

- Research infrastructures and partners joining the NAUTILOS project support free, open access (CC-BY) to data and metadata produced by partners' facilities and partners are committed to working towards the implementation of this principle;
- Data and metadata generated during the project are made available following an open access policy, without any restrictions and available for free to third parties (CC-BY);
- Appropriate controlled dictionaries (e.g. CF convention and SeaDataNet and NVS vocabularies, ISO8601) are recommended to be used for metadata description;

[1] https://www.force11.org/group/fairgroup/fairprinciples

- NAUTILOS data-products metadata will include a permanent identifier (doi)
- A metadata catalogue will be accessible on the NAUTILOS portal and data will be make available to facilitate NAUTILOS portfolio to projects and initiatives.

Contribution of data:

- The general responsibility for datasets that have been made available remains within the contributing institution/custodian/data originator;
- The quality assurance of data is the responsibility of the custodian/data originator and its declared doi;
- Data providers are requested to inform of any national policies that may place special conditions on the redistribution of data;
- Data licence is whenever possible CC-BY;
- Metadata shall be provided for each dataset following, as far as possible, agreed standards.

Use of data:

- Data interpretation is solely the responsibility of data user;
- Data sources shall be acknowledged, preferably using a formal citation (as indicated in the metadata), and where appropriate, the data originator shall be involved.

Data and metadata generated within NAUTILOS and which are stored at the originating institute/organization (which are the data owner-data provider and are responsible for the data, metadata and quality), are the same that are stored/made available with the NAUTILOS data portal (data assembly centre) and are the same that are made available to data integrator portals and initiatives (EMODnet, CMEMS, etc).
Whenever these integrators combine, integrate, elaborate, more generally process NAUTILOS data, it is recommended to acknowledge the data provenience (i.e. NAUTILOS) and the post-processing applied methods (for which NAUTILOS project is not responsible).
The NAUTILOS partners are not responsible for any use and misuse made by end-users.
This version will also report on data produced in the context of the project and non-sensitive data that can be made publicly available in open data repositories and registered at relevant catalogues.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition
AdriFOS	Adriatic Fisheries and Oceanography Observing System
ACDD	Attribute Convention for Data Discovery
AES	Advanced Encryption Standard
AtlantOS	Optimizing and Enhancing the Integrated Atlantic Ocean Observing System
BEIS	Department for Business, Energy & Industrial Strategy, UK
BODC	British Oceanographic Data Centre
CC	Creative Commons
CDI	Common Data Index
CF convention	Climate and Forecast convention
CMEMS	Copernicus Marine Environment Monitoring System
CMEMS INSTAC	In Situ Thematic Centre
CNR	Consiglio Nazionale delle Ricerche
CNR-IRBIM	Istituto per le Risorse Biologiche e le Biotecnologie Marine
DAC	Data Archive Centre
DATAMEQ	EuroGOOS Data Management, Exchange, and Quality Working Group

DB	Database
DEFRA	Department for Environment, Food & Rural Affairs
DMP	Data management policy
Doi	Digital object identifier
DOOS	Deep ocean observing system
EDMO	European Directory of Marine Organisations
EC	European Commission
EEA	European Environment Agency
EMODnet	European Marine Observation and Data Network
ENISA	European Union Agency for Cybersecurity
EOV	Essential Ocean Variables
ERDDAP	Environmental Research Division Data Access Program
EU	European Union
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EuroGOOS	European component of the Global Ocean Observing System of the Intergovernmental Oceanographic Commission of UNESCO
FAIR	Findability, accessibility, interoperability, and reusability
FOS	Fishery Observing System
FGDC	Federal Geographic Data Committee
G7	Group of Seven
G20	Group of Twenty
GDAC	Global Data Assembly Center
GES	Good Environment Status
GLODAP	Global Ocean Data Analysis Project
GOOS	Global Ocean Observing System
GRDP	General Data Protection Regulation
H2020	Horizon 2020 Framework Programme
HELCOM	Baltic Marine Environment Protection Commission
ICES	International Council for the Exploration of the Sea
INSPIRE	Infrastructure for Spatial Information in Europe
IOC/IODE	International Oceanographic Data and Information Exchange (IODE) of the Intergovernmental Oceanographic Commission
ISO	International Organization for Standardization
JERICO	Joint European Research Infrastructure of Coastal Observatories
JRC	Joint Research Center
MAP	Mediterranean Action Plan
MEDITS	An international bottom trawl survey in the Mediterranean
MSFD	Marine Strategy Framework Directive
NERC	Natural Environment Research Council
NetCDF	Network Common Data Form
NGOs	Non-governmental organizations
NODC	National Oceanographic Data Committee

NUG	NetCDF Users Guide
NVS	NERC Vocabulary Server
ODV	Ocean Data View 4
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
ORDP	Open Research Data Pilot document
OSPAR	Convention for the Protection of the Marine Environment of the North-East Atlantic
PANGAEA	Data Publisher for Earth & Environmental Science
SDN	SeaDataNet
SEANOE	Sea scientific open data edition
SOCAT	Surface Ocean CO2 ATlas
SOOS	Southern Ocean Observing System
SSL	Secure Sockets Layer
TAC	Thematic Assembly Centre
TSG ML	Technical Subgroup on Marine Litter
UN	United Nations
UNEP	UN Environment Programme
WAF	Web Accessible Folder
WCS	Web Coverage Service
WFS	Web Feature Service
WMTS	Web Map Tile Service
WMS	World Map Service
WoRMS	World Register of Marine Species

I. INTRODUCTION

Open ocean and deep-sea environments are a repository of valuable new knowledge on unexplored scientific phenomena, natural hazards and energy and sources opportunities. Due to the gap of long-term observation and technologically advanced systems, the research and exploitation of these environments is still at an immature and undeveloped stage.

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 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).

The goal of NAUTILOS is to fill in existing marine observation 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, in addition to micro- and nano-plastics, to improve our understanding of environmental change and anthropogenic impacts. Newly developed marine technologies will be integrated in different observing platforms and deployed by using novel approaches from shore to deep-sea deployments.

The fundamental aim of the project is to complement and expand current European observation tools and services, to obtain a collection of data at a much higher spatial and temporal resolution and coverage 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.

This document presents the project Data Management Plan (DMP) updated to month 36. In this version the main updates include the following:

- Association of data and metadata with new *dois*;
- Definition of standard vocabularies used for sharing the acquired data;
- Application of data management good-practice;

- Specification of metadata formats;
- Definition of the controlled vocabularies for parameters;
- Examples of standardised “Marine litter” data;
- Case studies added for Marine litter and Observing Systems;
- More detailed description of the Near-real-time data and delayed data;
- Detailed and final naming conventions based on catalogues;
- Examples of naming conventions;
- ERDDAP clarification, adoption and NAUTILOS data example;
- Added examples to interact and operate on the platform;
- Definition of the data availability from the backend infrastructure also through the NAUTILOS data portal (the front-end);
- Extended the definition of tools to access and operate on data;
- Definition and examples of the adopted NAUTILOS community on Zenodo;
- Specification on Open-Access for NAUTILOS validated data;
- Revised cv for the data management responsible person;
- Update on ethical aspects covered.

II. NAUTILOS PROJECT

NAUTILOS 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. This project will hence develop, integrate, validate and demonstrate new cutting-edge technologies with regards to sensors, interoperability and embedding skills. The development will always be guided by the **objectives of scalability, modularity, cost-effectiveness and open-source availability of data and software** products.

NAUTILOS is expected to collect, validate and process a huge amount of heterogeneous data that needs dedicated tools and services to favour integration and interoperability. Whenever possible, the developed data management infrastructure, tools and services will allow a data flow towards existing infrastructures and integrators globally accepted and used by the ocean observing community. Datasets acquired during the project through sensors and in-situ observation systems will be **made readily and freely available to these infrastructures and to the wider international ocean science community and other stakeholders**.

An on-line web user interface will provide the features to discover, access, retrieve sensors and platforms data, and will also represent the entry point for all the users (including citizen scientists) with an interest in validated environmental data collections. The interface will be designed considering specific dual requirements: an internal data storage and management area, and an externally visible and accessible area. Besides the classical marine data types, the project proposes to **acquire and manage new data** (e.g., digital images, micro-plastic observations etc.) whose harmonised data flow has yet to be designed and adopted at international level. The development is following models and schemes from the already existing infrastructures and propose itself as the champion/model for the establishment of these new data flows. This way, one key outcome of the project is to help adding and sharing more and better data and parameters. Behind the web interface, a standardisation and embedding process will make the data ready to dissemination and transfer to appropriate Thematic Assembly Centre (TACs) and data management infrastructures (e.g., EMODnet) in their respective accepted formats.

Data management is based on recommendations from key target groups and stakeholder initiatives to facilitate a fast adoption and availability of the produced data.
As also reported in other project reports (see D2.1) NAUTILOS is focusing on 17 instrumentation/tools (see Table 1 D2.1).

In this framework, NAUTILOS data management policy includes and clarifies the data life cycle, the roles on the ownership and custodianship of the data, as well as the recommendations for data flow and citation.

Following the structure of the Horizon 2020 DMP template[2], In this framework, the NAUTILOS DMP evolves during the lifespan of the project and successive versions present more details.

[2] C. Ramjoue and O. Marganne, "TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)," 13 October 2016. http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx.

III. DATA SUMMARY

Following the structure of the Horizon 2020 DMP template[2], in the following sections we present the NAUTILOS approaches to the key DMP indications.

1. WHAT IS THE PURPOSE OF THE DATA COLLECTION/GENERATION AND ITS RELATION TO THE OBJECTIVES OF THE PROJECT?

The purpose of NAUTILOS is to close the marine observation and modelling gaps for chemical, biological and deep ocean physics variables by means of next-generation cost-effective sensors and samplers, their deployment for large-scale demonstration in European seas and integration into observing platforms. This project therefore aims at complementing and expanding current European observation tools and services, to obtain a collection of data at a much higher spatio-temporal resolution and coverage, than currently available, and making the monitoring of the marine environment further available to both traditional and non-traditional data users.
In particular, the specific objectives of data collection are:

- to improve our understanding of environmental change and anthropogenic impacts related to aquaculture, fisheries, and plastic litter in coastal and shelf environments by means of improving current observing systems;
- to improve our understanding of open ocean and deep-sea environments;
- to improve the detection of plastic pollution to understand the input, distribution, and fate of plastics in European seas;
- to improve observing systems integrated to commercial activities such as fisheries, aquaculture, and ships of opportunity;
- to improve observing systems that utilise animal-borne instrumentation.

2. WHAT TYPES AND FORMATS OF DATA WILL THE PROJECT GENERATE/COLLECT?

The project deploys 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 physical, 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 nine MSFD descriptors.

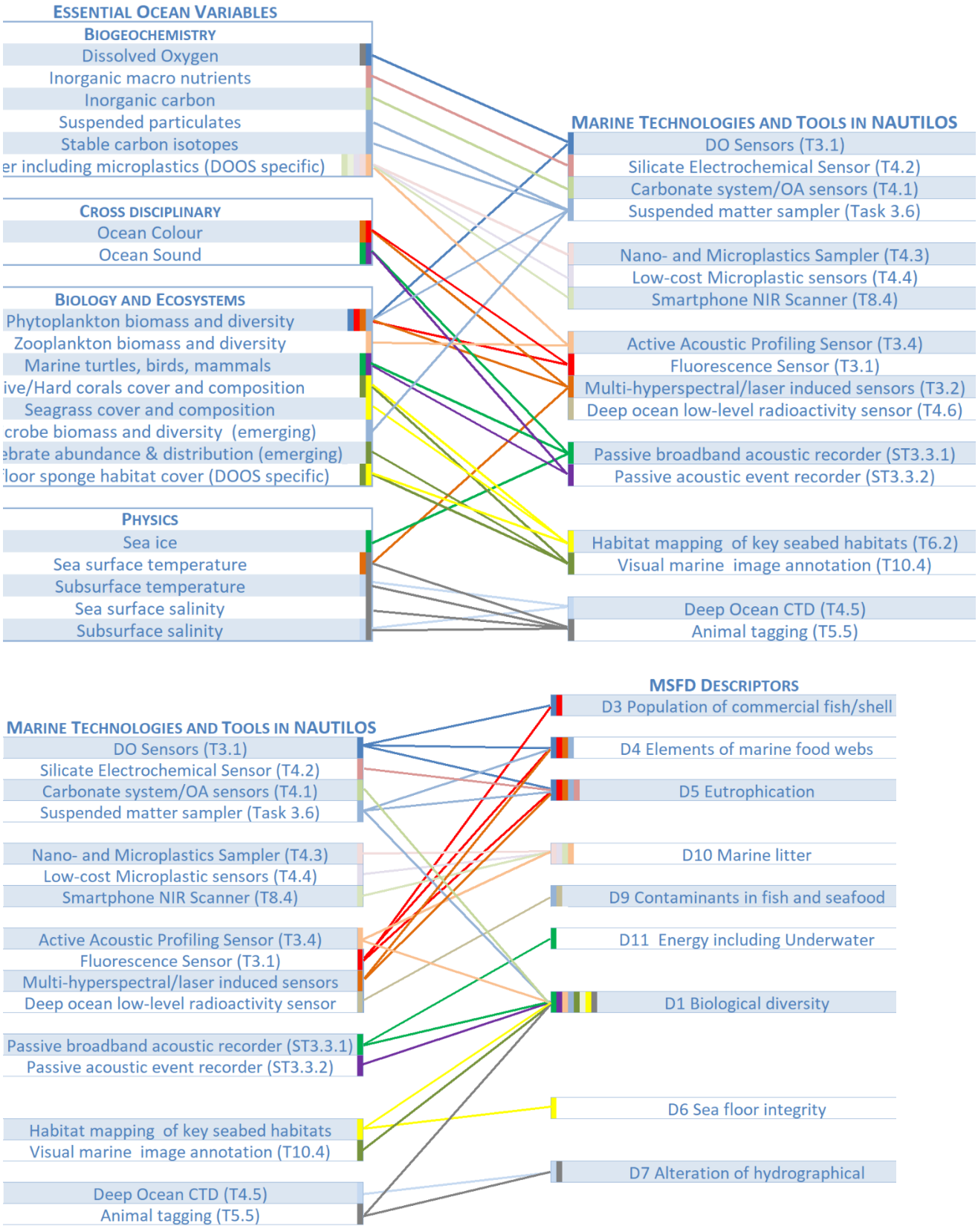

Figure 1 Environmental variables covered by NAUTILOS in terms of: Essential Ocean Variables (top); and MSFD descriptors (bottom) covered; updated following analysis performed in deliverable D2.1 (Figure 7).

There are two categories of data to be used within NAUTILOS:

- **internal project data**, i.e., data outputs from the project itself;
- **externally valuable environmental data**, i.e., real-world application data.

The data consists of a combination of numeric data (i.e., parameters measured by sensors), images and sounds from in situ fixed and moving platforms (time series, profiles, trajectories), and from model outputs (gridded data) that use collected data.

Those different types of data require different management methodology, storage capacities and standards.

While the data flow for classical ocean physical parameters such as temperature and salinity is well-defined[3], some of the NAUTILOS variables, such as digital images, microplastic observations, and acoustic data, are part of new data flows and while the endpoint may be already defined, the in-between data standards are still under discussion.

"Marine Litter" as a Case Study

Marine Litter has been added to the EMODnet Chemistry scope since 2017. It is an important subject on the international political agendas such as of G7 and G20. It is very relevant for the MSFD agenda and is managed under the descriptor D10. This aims to provide instruments to assess, monitor, set targets and finally reach a good environmental status (GES) with regard to marine litter. GES should be achieved only when “properties and quantities of marine litter do not cause harm to the coastal and marine environment”.
To this end EMODnet Chemistry has developed products for these three main categories:

- Beach litter (nets, bottles etc.)
- Seafloor Litter (i.e. litter collected by fish trawl surveys)
- Micro-litter (micro plastics)

Starting from the outcomes of already ongoing initiatives (Technical Support Group – Marine Litter (TSG ML), JRC Project on Marine Litter baselines, Regional Sea Conventions (OSPAR, HELCOM, UNEP/MAP, BSCS), ICES, MEDITS, etc.), EMODnet Chemistry implemented two main databases - one for beach litter, modelled after the OSPAR-MCS approach, and one for seafloor litter, modelled after the ICES-DATRAS approach – and they collect a description of the detected elements, i.e. standardized description of the sampled element by using common terms from a standardized vocabulary.

[3] https://eurogoos.eu/data-management-exchange-quality-working-group-data-meq/

In other terms, whatever is the methodology to collect the sample (manual annotation, taking a picture, taking a sample and processing it in the lab, etc.) the outcome of the procedure is a collection of information describing the litter.

Size	Record size of each item. Minimum resolution is to allocate in to bin sizes of 100 µm
Type	Plastic fragments, pellets, filaments, plastic films, foamed plastic, granules, and styrofoam
Shape	For pellets: cylindrical, disks, flat, ovoid, spheruloids; For fragments: rounded, subrounded, subangular, angular; For general- irregular, elongated, degraded, rough, and broken edges
Colour	Transparent, crystalline, white, clear-white-cream, red, orange, blue, opaque, black, grey, brown, green, pink, tan, yellow

CATEGORIES FOR MICROPARTICLES	
Material	**Description**
Plastic	Plastic fragments rounded
	Plastic fragments subrounded
	Plastic fragments subangular
	Plastic fragments angular
	cylindrical pellets
	disks pellets
	flat pellets
	ovoid pellets
	spheruloids pellets
	filaments
	plastic films
	foamed plastic
	granules
	styrofoam
Other	Other (glass, metal, tar)

Figure 2 Categories to describe microplastics appearance (extracted from Guidance on Monitoring of Marine Litter in European Seas, MSFD Technical Subgroup on Marine Litter, 2013).

This information is collected in a standard metadata format, i.e. Common Data Index (CDI)[4] and ingested in the EMODnet Chemistry DB.

The scope of NAUTILOS DMP is to identify and describe these endpoints, to support NAUTILOS partners to collect and describe the relevant information that may be provided (or made available) to integrating infrastructure.

The final goal of NAUTILOS is to organize its data and make it accessible for relevant stakeholder by applying FAIR principles and exploiting machine-to-machine interoperability.

For the well-established parameters, NAUTILOS data management will adopt and implement standards from existing infrastructures (DACs and Global DACs) and integrators (e.g., EMODnet, CMEMS, JERICO[5], ICES[6], OGC[7], DarwinCore[8], etc) and once data is ready in the

[4] https://www.emodnet-chemistry.eu/repository/Proposal-EMODnet-TG-ML-Micro-Litter-Data-Gathering-03062020.pdf
[5] https://www.jerico-ri.eu/
[6] https://www.ices.dk/Pages/default.aspx
[7] https://www.ogc.org/
[8] https://dwc.tdwg.org/

NAUTILOS back-end data infrastructure, stakeholder will be notified to start their uptaking process.

A Case Study on Ocean and Meteorological Data Collection within an Opportunity Observing System (FOS)

Since 2021, the Ancona section of CNR-IRBIM is running the AdriFOS initiative: fishing vessels equipped with an integrated system for the collection of information on catches, position of the fishing operation, depth and water temperature during the haul, are producing a great amount of data for oceanographic (and biodiversity) purpose. The NAUTILOS DMP design how this data has to be organized in the NAUTILOS backend system and define the procedure to inform NAUTILOS stakeholder that these data are available, more specifically, the NAUTILOS Data Manager is recommended to inform NAUTILOS stakeholder or initiate stakeholder ingestion procedure to let this data be consumed, considering that this data can be operationally shared to EMODnet Physics, CMEMS INSTAC, etc.. In what concerns the EMODnet Ingestion system[9], the NAUTILOS Data Manager is recommended to activate this EMODnet Ingestion procedure and follow up towards the integration steps.

Regarding NAUTILOS project activities, a series of documents and reports are planned coping with data management and its strictly connected issues, as agreed in the Annex 1 to the Grant Agreement. The list of documents follows in Table 1.

Table 1. List of data management documents in NAUTILOS

Document	Description with respect to Data Management and Nature
D1.3 Data Management Plan	Public ORDP: Open Research Data Pilot document. Submitted at M6
D1.10 Data Management Plan – 1st periodic report update	Update of DMP at M18. Public ORDP: Open Research Data Pilot document. Submitted and approved at M18.
D1.11 Data Management Plan – 2nd periodic report update	This document Update of DMP at M36. Public ORDP: Open Research Data Pilot document.
Final Data Management Plan	Finalization of the DMP at the end of the project. Public ORDP: Open Research Data Pilot document.
D8.3 Data Management Workflow	Description of common methods for parameter-platform management. Public report submitted and approved at M12.
D8.4 Design of Thematic Assembly Center for innovative parameters	Description of how to set up a dedicated NAUTILOS assembly centre for new parameters.

[9] https://www.emodnet-ingestion.eu/operational-data

	Other: design of software platform submitted and approved at M12.
D9.5 KPI definition for the NAUTILOS data management and dissemination infrastructure	Description of the methodology for assessing production of new valuable data. Public report submitted at the end of 2nd year.
D9.6 KPI assessment 1	Document for tracking impact of NAUTILOS in terms of data management and dissemination. Public report submitted at the end of 3rd year.
D9.7 KPI Assessment 2	Document for tracking impact of NAUTILOS in terms of data management and dissemination, final update. Public report due at the end of the project.

2.1. Common data formats

The primary data format for the NAUTILOS data distribution is going to be the OceanSites netCDF-4 classic model[10]. NetCDF (Network Common Data Form) is a set of software libraries and machine - independent data formats that is the international standard for common data and it is the one adopted by all key European and international ocean data management infrastructures (Global DAC, CMEMS, EMODnet, SeaDataNet, etc.).

The recommended implementation of NetCDF is based on the community-supported CF Convention, which provides a definitive description of the data in each variable, and the spatial and temporal properties of the data. The used version is CF-1.6 and it shall be identified in the 'Conventions' attribute. Any relevant metadata should be included whether it is part of the standard or not.

To fulfil its objectives and facilitate fast integration into international ocean data management infrastructures (e.g., GDAC, EMODnet, etc), NAUTILOS also adds some requirements to the CF-1.6 standard:

- Where time is specified as a string, the 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;
- Global attributes from Unidata's NetCDF Attribute Convention for Data Discovery (ACDD) are implemented;
- INSPIRE directive compliance is recommended;
- Variable names (short names) from SeaDataNet P02 controlled vocabulary[11] are recommended;

[10] http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf
[11] https://vocab.seadatanet.org/v_bodc_vocab_v2/vocab_relations.asp?lib=P02

- Institution codes: EDMO (European Directory of Marine Organisations)[12].

2.2. Metadata and Global attributes

The global attribute section of a NetCDF file describes the contents of the file overall, and allows for data discovery. All fields should be human-readable and use units that are easy to understand. Global attribute names are case sensitive.

The European common data and metadata model for real-time data divides global attributes to be adopted for data in three categories: Mandatory Attributes, Recommended Attributes and Suggested Attributes.

The Mandatory Attributes (M) include attributes necessary to comply with CF-1.6 and OceanSITES conventions. The Recommended Attributes (R) include attributes necessary to comply with INSPIRE and Unidata Dataset Discovery conventions.

The Suggested Attributes (i.e., the others) include attributes that can be relevant in describing the data, whether it is part of the standard or not. All these attributes should be used and contain meaningful information, unless there are technical reasons making this impossible.

Attributes are organized by function: Discovery and Identification, Geo-spatial- temporal, Conventions used, Publication information, and Provenance. Attributes that are part of the Attribute Convention for Data Discovery (ACDD) or Climate and Forecast (CF) standard, or that appear in the NetCDF Users Guide (NUG) are so indicated, as are those that are used by GDAC inventory software.

The application of thisdata management good practice speeds up the compilation and publication of the metadata (i.e. the Common Data Index – CDI) for the validated datasets under the SeaDataNet network of National Oceanographic Data Centres. CDI is assigned the National Oceanographic Data Centre that has applied the SDN procedures for long term stewardship of the data.

The CDI metadata format is a marine profile of the ISO19115/ISO19139 metadata standard and it is supported for many metadata tags by the SeaDataNet common vocabularies and directories.

Adopting the above described data management good practice for metadata speeds up and facilitates the activity of NODCs when checking and ingesting new datasets into SDN validated research quality ocean data DB.

2.3. NAUTILOS Recommendations for interoperability in a nutshell

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

[12] https://edmo.seadatanet.org/

- (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) P01/P02/P09
    - Units: SDN::P06
    - Institution codes: EDMO (European Directory of Marine Organisations)
    - Country code: ISO 3166
- Data publishing service: ERDDAP + GeoServer
- Data catalogue service: GeoNetwork

Table 2. Controlled Vocabularies for Parameters

ID	Title	Version	version date	Description	Governance	new terms request - link
P01	BODC Parameter Usage Vocabulary	1170	04/09/23	Terms built using the BODC parameter semantic model designed to describe individual measured phenomena. May be used to mark up sets of data such as a NetCDF array or spreadsheet column. Units must be specified when using a P01 code. The P06 unit that is linked to individual P01 in the NVS is the one used in BODC's systems but external users can use any appropriate units.	British Oceanographic Data Centre	https://github.com/nvs-vocabs/P01
P02	SeaDataNet Parameter Discovery Vocabulary	126	31/08/23	Terms describing fine-grained related groups of measurement phenomena designed to be used in dataset discovery interfaces.	SeaDataNet	https://github.com/nvs-vocabs/P02
P09	MEDATLAS Parameter Usage Vocabulary	74	20/04/23	Terms under the content governance of SISMER used to describe measured phenomena within the MEDATLAS project.	Systèmes d'Informations Scientifiques pour la Mer	

P06	BODC-approved data storage units	141	05/09/23	Terms approved for use by BODC to describe the measurement units for data held in its repositories.	British Oceanographic Data Centre	https://github.com/nvs-vocabs/P06

Moreover, to facilitate the uptake of the Marine Litter data into the EMODnet Chemistry DB, NAUTILOS partners may use P36 (instead of P02) more specifically SDN::P36::MRNLTTR that is grouping the P02 terms on marine and beach litter.

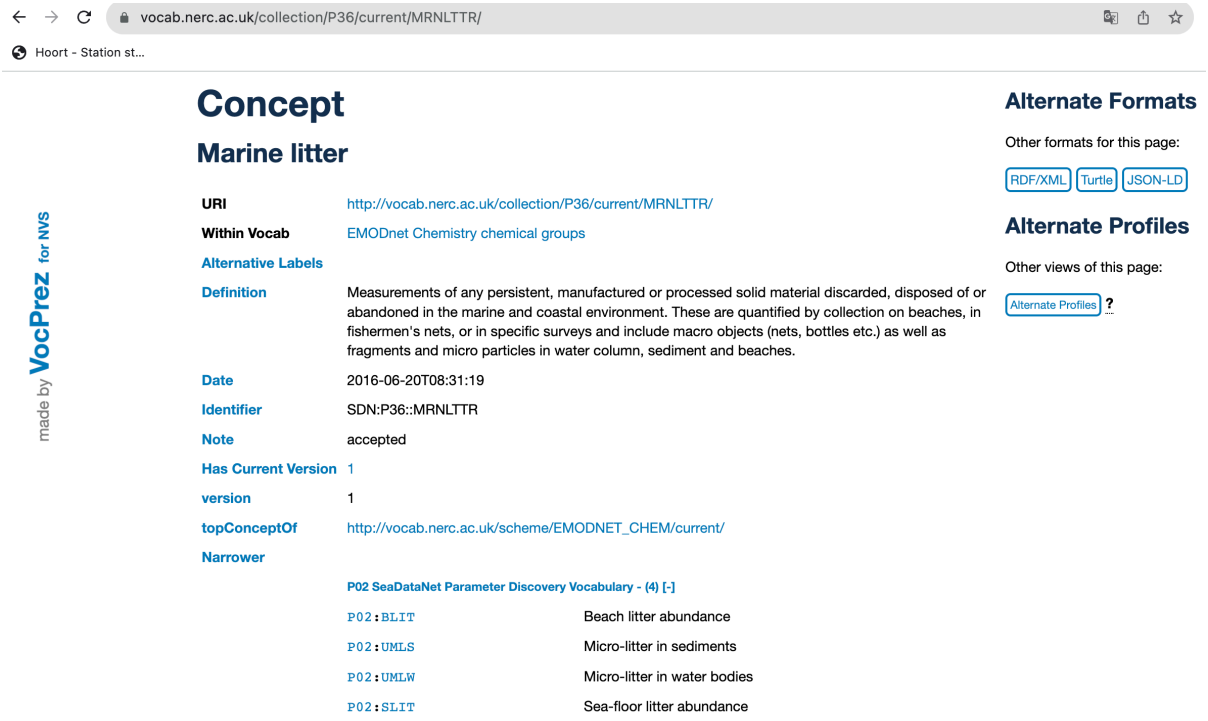

Figure 3- SDN::P36::MRNLTTR

3. WILL YOU RE-USE ANY EXISTING DATA AND HOW?

NAUTILOS is using relevant publicly available oceanographic dataset, such as those made available by key European data integration infrastructures:

- For In Situ data the primary sources are: CORIOLIS, CMEMS INSTAC, EMODnet, SeaDataNet.
- For Remote Sensing data the sources are: CMEMS, NASA (Aqua – MODIS), EUMETSAT (MetOP)

To note that, according EUMETSAT license agreement, data can be used by NAUTILOS partners for internal use (e.g. validation) but cannot be redistributed.
This data is going to be available to NAUTILOS partner by NAUTILOS data infrastructure – private area – under authentication and restriction systems (see also D8.3).

4. WHAT IS THE ORIGIN OF THE DATA?

To better understand the oceans, ocean observing technologies must be able to assess the spatial and temporal heterogeneity in the ocean with regards to physical processes, distribution of elements, ocean productivity, microbial to megafaunal biodiversity, and anthropogenic impacts related to fossil fuel emissions (i.e., ocean acidification), chemical pollutants, and litter/(micro)plastics. To achieve this, the ocean observing community must develop widely distributed observing systems equipped with low-cost and modular sensors and samplers on a variety of observing platforms.
NAUTILOS data is originated by the NAUTILOS sensors (oxygen, temperature, salinity, fluorescence, conductivity, light etc) that are mounted and carried by operating platforms (e.g., Ships of Opportunity, Fishing vessels, AUVs, tagged sea mammals etc).
Details on sensors specification are described in NAUTILOS WP3-WP4 Deliverables.

5. What is the expected size of the data? Who might it be useful to ('Data Utility')?

The expected size of the data is going to be in the range of terabytes.
The collected data will be made available to the wider community of ocean data users. NAUTILOS data are useful to **policy and decision makers**, including the European Commission, Parliaments, Members States officials, UN bodies, HELCOM and OSPAR commissions, EEA, supporting agencies of member state legislators, state agencies, governmental bodies, national funding agencies (i.e., DEFRA, BEIS in the UK). Many observation and monitoring programmes inform policies designed to enable the protection of the global oceans. The observation and monitoring technologies developed within the project can significantly contribute to policy aimed at promoting the good environmental status, conservation and protection of marine ecosystems. Key policy makers can benefit from a close collaboration with the project, they will be informed about the project's results, and will be actively engaged in providing feedback on whether project outcomes address current limitations, match the needs of those making decisions on policies affecting marine environment and answer future needs.

Other users are **commercial and industrial activities**, such as the fishing industry, aquaculture operators, offshore energy industry (oil and gas exploration, wind and tidal generation), seabed extractive activities, the tourism and recreation sector, marine biotechnology and bioprospecting, telecommunications, coastal protection, defence, search and rescue. Blue economy is a major contributor to the European economy, and the socio-economic benefits provided by the ocean are reliant on observations, measurements, and forecasts.

The **ocean research community** is definitely also the target of this project, as will greatly benefit from the data generated for carrying out a variety of **projects in the areas of marine and earth observation**.
NAUTILOS is also directed to ocean-focused **NGOs and citizen scientists**, as they can share the knowledge, raise awareness, participate in project campaigns and field work.

Finally, the data generated by NAUTILOS have the potential of supporting the generation of several downstream services and applications targeting the general public.

IV. FAIR DATA

NAUTILOS' general data management policy that is presented in the following sections has been developed in accordance with Horizon 2020 FAIR principles[13], open data requirements and implementation guidelines.

It applies mainly to new results that will be produced in NAUTILOS that are to be made available by the project consortium as open source, open science and open data.

1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

i. Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g., persistent and unique identifiers such as Digital Object Identifiers)?

The data produced and used in the project will be discoverable with metadata and identifiable by means of a standard identification mechanism.

The collection and management of marine data metadata is a complex stepwise process that includes selection of the sensors, their configuration, the deployment at sea by means of a hosting platform, collection of the observations that may be transmitted to a shore-located receiving station or recovered at the end of the mission. Near real time transmission or delayed mode recovery largely depends on the platform in use e.g. a fixed coastal station/moored buoy may operate in real-time (where data is transmitted by telecommunication systems or via cables), an autonomous unmanned vehicle like a glider may operate in delayed mode (and data are downloaded from the AUV once the system is recorded), a ship based sensor may operate in real time (e.g. temperature) or delayed mode (chemical concentration, biota needs a further sample processing). Each step should be documented and adopt best practices and standards. As described in the previous paragraphs NAUTILOS acts as an integrator entity: NAUTILOS sensors and platforms are managed by NAUTILOS partners to collect data, these data are delivered back-to-back with metadata that adopts common standards. More specifically NAUTILOS designed a minimum number of metadata elements to provide users with information identifying a collection of files as a thematic/coherent dataset. This also includes a naming convention (i.e. test, nrt, etc. dataset name prefix) for datasets published into the NAUTILOS data catalogue (i.e. ERDDAP)

NAUTILOS data catalogue supports the search through those collections using keywords and spatio-temporal coordinates and will provide information on or links to the processing history of the observations (i.e., source, version, quality assessment and control, sensors).

The adoption of ISO standards and the use of shared controlled vocabularies are a key prerequisite towards consistency and this data integrator and data mediator role. ISO 19115

[13] https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Standard[14] requires a basic minimum number of metadata elements that are essential for the data presentation:

- Dataset or dataset series on specific challenges ('what'),
- Geographic bounding box ('where'),
- Temporal extent ('when'),
- Contact point to learn more about or order the dataset ('who').

The key references for cataloguing the information used in NAUTILOS are:

- ISO 8601 Representation of date and time,
- SeaDataNet NVS P0x description of parameters,
- Climate and Forecasting conventions for parameters standard names,
- WGS84 for Datum.

In line with GOOS recommendations on tools to make data FAIR, to facilitate the data harmonisation and to operate as an integrator and data translator for facilitating data use and interoperability, NAUTILOS adopted ERDDAP[15] as the core solution for data management (see D8.3 for detailed description on data management back-end infrastructure).

The NAUTILOS ERDDAP is accessible at https://data-nautilos-h2020.eu/erddap/

Each dataset is enriched with metadata: as an example, the metadata of the Fishing Vessel Campaigns in the Adriatic Sea dataset are available at

https://data-nautilos-h2020.eu/erddap/info/AdriFOOS_profiles_2012-2020/index.html

Table 3. AdriFOOS dataset metadata (to note that the system lists the attibute in alphabetical order)

Row Type	Variable Name	Attribute Name	Data Type	Value
attribute	NC_GLOBAL	cdm_data_type	String	Point
attribute	NC_GLOBAL	Citation	String	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.
attribute	NC_GLOBAL	citation	String	Penna Pierluigi, Belardinelli Andrea, Croci Camilla Sofia, Domenichetti Filippo, Martinelli Michela.
attribute	NC_GLOBAL	Conventions	String	COARDS, CF-1.6, ACDD-1.3
attribute	NC_GLOBAL	DOI	String	https://doi.org/10.17882/73008
attribute	NC_GLOBAL	Easternmost_Easting	double	18.60072
attribute	NC_GLOBAL	EDMO	double	5060

[14] https://www.iso.org/standard/53798.html

[15] https://www.ncei.noaa.gov/erddap/index.html

attribute	NC_GLOBAL	featureType	String	Point
attribute	NC_GLOBAL	geospatial_lat_max	double	44.97817
attribute	NC_GLOBAL	geospatial_lat_min	double	40.39051
attribute	NC_GLOBAL	geospatial_lat_units	String	degrees_north
attribute	NC_GLOBAL	geospatial_lon_max	double	18.60072
attribute	NC_GLOBAL	geospatial_lon_min	double	12.4062
attribute	NC_GLOBAL	geospatial_lon_units	String	degrees_east
attribute	NC_GLOBAL	infoUrl	String	https://www.seanoe.org/data/00618/73008/
attribute	NC_GLOBAL	institution	String	CNR IRBIM
attribute	NC_GLOBAL	keywords	String	cruise, data, dbar, latitude, longitude, ocean, pressure, QV_ODV_SAMPLE, QV_SEADATANET, QV_SEADATANET_1, source, station, temperature, time, type.
attribute	NC_GLOBAL	license	String	CC-BY 4.0
attribute	NC_GLOBAL	Northernmost_Northing	double	44.97817
attribute	NC_GLOBAL	sourceUrl	String	(local files)
attribute	NC_GLOBAL	Southernmost_Northing	double	40.39051
attribute	NC_GLOBAL	standard_name_vocabulary	String	CF Standard Name Table v70
attribute	NC_GLOBAL	subsetVariables	String	Type, Cruise, Station
attribute	NC_GLOBAL	summary	String	CNR-IRBIM implemented the "AdriFOOS" observational system, by installing the FOOS on some commercial fishing boats operating in the Adriatic Sea. Since then the datacenter based in Ancona receives daily data sets of environmental parameters collected along the water column and close to the sea bottom (eg. temperature, salinity, etc.), together with GPS haul tracks, catch amounts per haul, target species sizes and weather information.
attribute	NC_GLOBAL	time_coverage_end	String	2020-02-26T11:43:25Z
attribute	NC_GLOBAL	time_coverage_start	String	2012-11-26T05:12:52Z
attribute	NC_GLOBAL	title	String	Adri FOOS Temperature profiles 2012-2020.
attribute	NC_GLOBAL	Westernmost_Easting	double	12.4062
variable	Cruise		String	
attribute	Cruise	long_name	String	Cruise
variable	Station		short	
attribute	Station	_FillValue	short	32767
attribute	Station	actual_range	short	1, 154
attribute	Station	long_name	String	Station
variable	Type		String	
attribute	Type	long_name	String	Type
variable	time		double	
attribute	time	_CoordinateAxisType	String	Time
attribute	time	actual_range	double	1.353906772E9, 1.582717405E9
attribute	time	axis	String	T
attribute	time	ioos_category	String	Time
attribute	time	long_name	String	YYYY-MM-DD THH:MM:SS.SSS
attribute	time	source_name	String	yyyy-mm-ddThh:mm:ss.sss
attribute	time	standard_name	String	time
attribute	time	time_origin	String	01-JAN-1970 00:00:00

attribute	time	time_precision	String	1970-01-01T00:00:00Z
attribute	time	units	String	seconds since 1970-01-01T00:00:00Z
variable	longitude		float	
attribute	longitude	_CoordinateAxisType	String	Lon
attribute	longitude	actual_range	float	12.4062, 18.60072
attribute	longitude	axis	String	X
attribute	longitude	colorBarMaximum	double	180.0
attribute	longitude	colorBarMinimum	double	-180.0
attribute	longitude	ioos_category	String	Location
attribute	longitude	long_name	String	Longitude
attribute	longitude	source_name	String	Longitude [degrees_east]
attribute	longitude	standard_name	String	longitude
attribute	longitude	units	String	degrees_east
variable	latitude		float	
attribute	latitude	_CoordinateAxisType	String	Lat
attribute	latitude	actual_range	float	40.39051, 44.97817
attribute	latitude	axis	String	Y
attribute	latitude	colorBarMaximum	double	90.0
attribute	latitude	colorBarMinimum	double	-90.0
attribute	latitude	ioos_category	String	Location
attribute	latitude	long_name	String	Latitude
attribute	latitude	source_name	String	Latitude [degrees_north]
attribute	latitude	standard_name	String	latitude
attribute	latitude	units	String	degrees_north
variable	SN		String	
attribute	SN	long_name	String	SN
variable	Pressure		double	
attribute	Pressure	_FillValue	double	NaN
attribute	Pressure	actual_range	double	-1.93, 288.089996
attribute	Pressure	long_name	String	Pressure
attribute	Pressure	units	String	dbar
variable	QV_SEADATANET		byte	
attribute	QV_SEADATANET	_FillValue	byte	127
attribute	QV_SEADATANET	actual_range	byte	0, 9
attribute	QV_SEADATANET	long_name	String	QV:SEADATANET
variable	Temperature		double	
attribute	Temperature	_FillValue	double	NaN
attribute	Temperature	actual_range	double	6.29, 28.93
attribute	Temperature	long_name	String	Temperature
attribute	Temperature	SDN	String	SDN:P01::TEMPPR01
attribute	Temperature	units	String	degrees_C
variable	QV_SEADATANET_1		byte	
attribute	QV_SEADATANET_1	_FillValue	byte	127
attribute	QV_SEADATANET_1	actual_range	byte	0, 9
attribute	QV_SEADATANET_1	long_name	String	QV:SEADATANET.1

variable	QV_ODV_SAMPLE		byte	
attribute	QV_ODV_SAMPLE	_FillValue	byte	127
attribute	QV_ODV_SAMPLE	actual_range	byte	1, 1
attribute	QV_ODV_SAMPLE	long_name	String	QV:ODV:SAMPLE

ERDDAP supports both human interaction (e.g., OPeNDAP requests) and machine-to-machine interoperability. Moreover, 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 these formats.

Considering the example on the Fishing Vessels Campaigns in the Adriatic Sea, these features are explorable at the dataset page: https://shorturl.at/kpANY, here the user can fine tune the selection and define the query to download a chuck of the dataset in a given format (e.g. csv):

https://data-nautilos-h2020.eu/erddap/tabledap/AdriFOOS_profiles_2012-2020.csv?Cruise%2CStation%2CType%2Ctime%2Clongitude%2Clatitude%2CTemperature%2CQV_SEADATANET_1&Cruise=%22AN-01%22&time%3E=2012-01-01T00%3A00%3A00Z&time%3C=2021-12-31T00%3A00%3A00Z

Figure 4. Script to download (in csv) all the data collected by the cruise «AN-01» between 2012 and 2021

Notably, ERDDAP implements FGDC Web Accessible Folder (WAF) with FGDC-STD-001-1998 and ISO 19115 WAF with ISO 19115-2/19139. Besides ERDDAP, NAUTILOS hosts a GeoServer that implements several Open Geospatial Consortium protocols including Web Map Service (WMS), Web Feature Service (WFS), Web Coverage Service (WCS) and Web Map Tile Service (WMTS) and that was lately updated with the INSPIRE module. To extend further the users and uses of its products, NAUTILOS also implements web APIs and widgets.

ii. What naming conventions do you follow?

Whenever it is not possible to apply and harmonized approved naming convention (e.g., from EuroGOOS DATAMEQ), NAUTILOS will apply a human readable/understandable naming, e.g. if a dataset comes from preliminary tests (hence it is for internal use and not for stakeholder consumption) it clearly states this status in the dataset name.

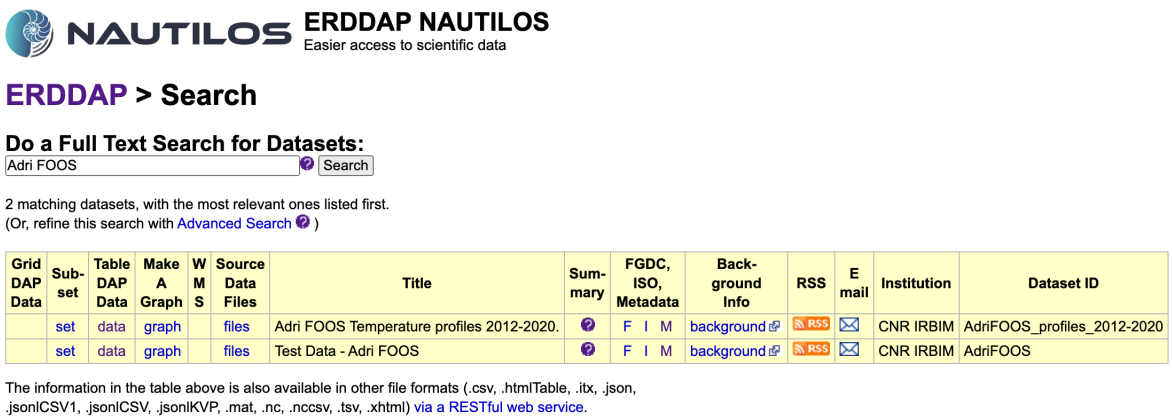

Figure 5. example of naming convention

iii. Will search keywords be provided that optimize possibilities for re-use?

Yes, search keywords will be provided based on the metadata and naming conventions.

iv. Do you provide clear version numbers?

All NAUTILOS products will be clearly labelled, as well as the version/type is declared (see also Figure 5)

v. What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

NAUTILOS follows the recommendations set up by the EuroGOOS DATAMEQ, developed under CMEMS, EMODnet Physics and SeaDataNet/SeaDataCloud and further extended at international level under the AtlantOS project[16] and EuroSEA project[17].

Metadata used by the networks for parameters should be "mappable" on standard vocabularies existing and EU (SeaDataNet vocabularies) or international (CF or WoRMS for Taxa). More specifically, metadata are based on P01-P02 (parameter), P07 (CF variable), P06 (units) from SeaDataNet controlled vocabularies managed by NERC/BODC (Vocabulary Server (version 2.0). Other relevant vocabularies are P36 and A05, mapping parameters In EMODnet Chemistry and ECV-EOV respectively.

Data provenance should be identifiable in the metadata (version, provider-long name, EDMO), as well as quality checks and flags (whenever applied/applicable)

[16] https://www.atlantos-h2020.eu/

[17] www.eurosea.eu

2. MAKING DATA OPENLY ACCESSIBLE

i. Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Validated data that are relevant for NAUTILOS stakeholders will be freely accessible to community as soon as possible. This may be instantaneous or may take a short period to make sure that the required processing and quality control have been performed, as well as, allowing partners some time to report results in scientific journals.

Data will be available on the NAUTILOS web portal and via its backend infrastructure (e.g. https://data-nautilos-h2020.eu/erddap/index.html)

ii. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

This will be reported periodically.

iii. How will the data be made accessible (e.g., by deposition in a repository)?

Data and products will be integrated into the NAUTILOS data infrastructure and NAUTILOS ERDDAP to facilitate the access, reuse and further improvements of these results/products. As described in previous sections, the NAUTILOS infrastructure implements the most recent dissemination catalogues and technologies and links and delivers data into the most relevant ocean data management infrastructures and programs.

iv. What methods or software tools are needed to access the data?

Data are open and freely available and can be viewed and used by using well-known software tools. NAUTILOS is also developing Colab/jupyler notebooks to offer step-by-step examples to consume NAUTILOS data. These scripts are going to be made available in the web portal.

Install all the packeges

```
!pip install seaborn
!pip install matplotlib
```

Import them

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns  # Import seaborn for color palette
```

Plot from ERDDAP

```
# Load the CSV data, skipping the second row
url = "https://data-nautilos-h2020.eu/erddap/tabledap/osc_plastic_litter.csv?campaign_id%2Cplastic_type%2Cquantity%2Cuid%2Ctim
df = pd.read_csv(url, skiprows=[1])

# Filter out rows where plastic_type is "Other"
df = df[df['plastic_type'] != 'Other']

# Group the data by plastic_type and calculate the sum of quantities
plastic_type_quantity = df.groupby('plastic_type')['quantity'].sum().reset_index()

# Create a color palette with a unique color for each plastic type
colors = sns.color_palette('pastel', n_colors=len(plastic_type_quantity))

# Create a bar graph with individual colors for each plastic type
plt.figure(figsize=(10, 6))
plt.bar(plastic_type_quantity['plastic_type'], plastic_type_quantity['quantity'], color=colors)
plt.xlabel('Plastic Type')
plt.ylabel('Quantity')
plt.title('Plastic Type vs. Quantity')
plt.xticks(rotation=90)
plt.tight_layout()

# Show the graph
plt.show()
```

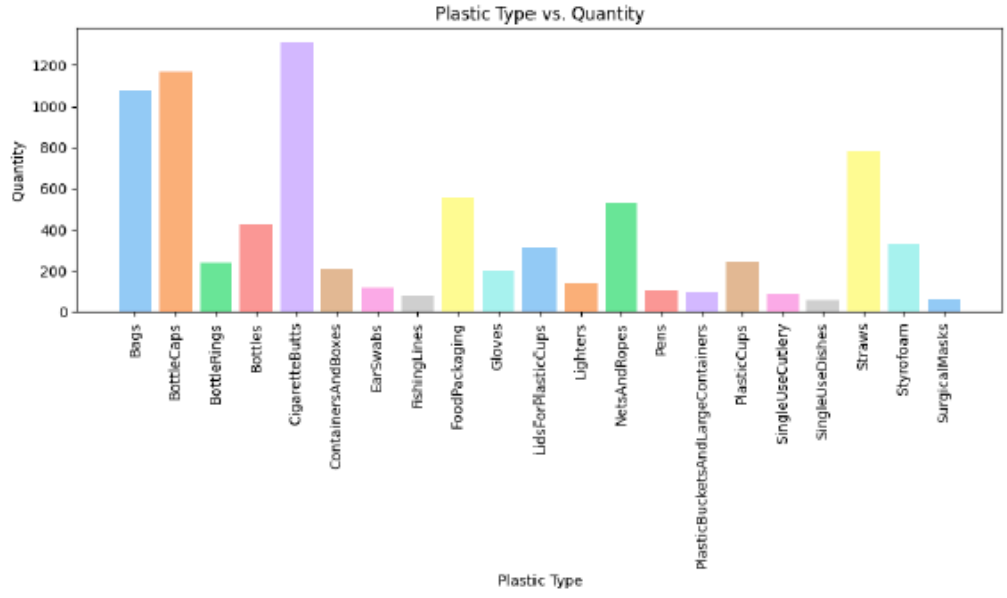

Figure 6. Colab notebook example

v. Is documentation about the software needed to access the data included?

Although the preference will always be open-source software, links to any needed software and documentation will be provided.

vi. Is it possible to include the relevant software (e.g., in open source code)?

All the needed software for accessing the data and products will be provided through the NAUTILOS portal, and open-source code will be available through typical defined Open Access repositories.

vii. Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

As described in the previous sections, data, metadata and documentation will be disclosed and deposited to key European ocean data infrastructures and repositories such us EMODnet, CMEMS, SDN and the NODC network, SEANOE. Other open repositories and initiatives such ICES, PANGAEA, SOOS, SOCAT, etc. will also be used.

Moreover, to facilitate access and visibility of public project documents, Zenodo repository has been adopted.

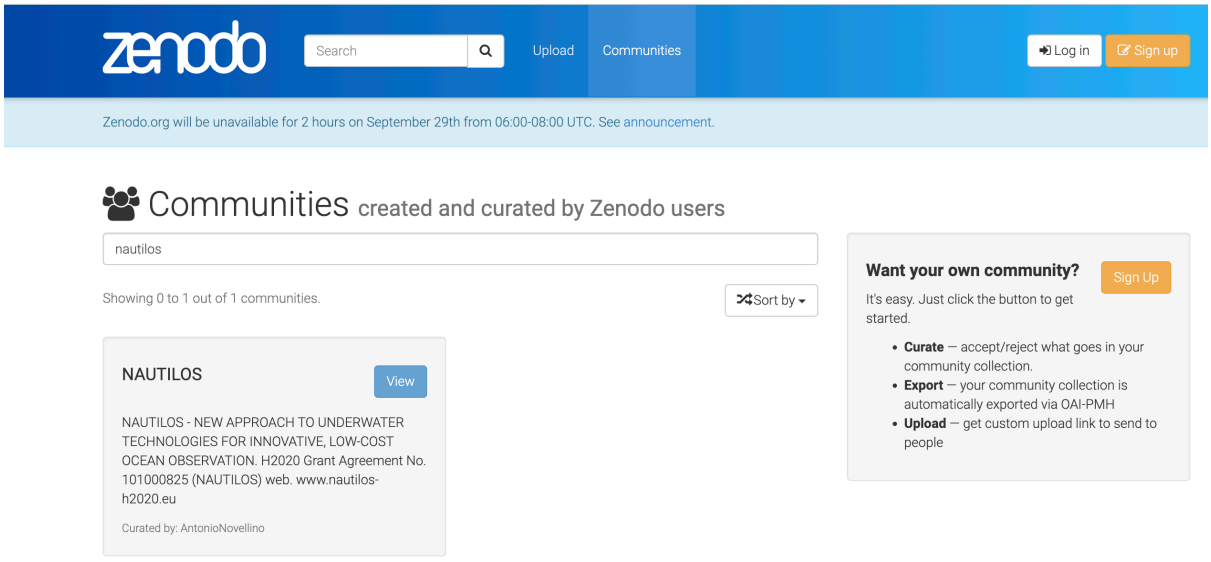

Figure 7. NAUTILOS community on Zenodo

https://zenodo.org/communities/nautilos-h2020/?page=1&size=20

viii. Have you explored appropriate arrangements with the identified repository?

Project partners are involved in or are working with systems and infrastructures (see point above) and some arrangements are already in place.

ix. If there are restrictions on use, how will access be provided?

NAUTILOS validated data are CC-BY and fully accessible with no restrictions.

x. *Is there a need for a data access committee?*

We do not see a need for a data access committee because all regulations are unambiguous.

xi. *Are there well described conditions for access (i.e., a machine-readable license)?*

Machine-to-machine is based on the most recent technologies, the premise is a CC BY-SA or CC BY-NC license.

xii. *How will the identity of the person accessing the data be ascertained?*

Once data is published, it is public and it is accessible without any restriction nor authentication needed.

NAUTILOS is monitoring users use (in compliance with GRDP) as described in D9.5. Another means for collecting feedback is by survey: specific templates for various usages of informed consent and all the procedures defined within NAUTILOS have been provided and specified within the deliverables D13.1 and D13.2, pertinent to the Ethics Work Package 13. Updated informed consent forms, analysed within the NAUTILOS Ethical Advisory Board, have been developed and submitted as annexes of the deliverable D13.7. In the deliverable different versions of the forms are present in different languages which can be used by different NAUTILOS partners in different countries.

3. MAKING DATA INTEROPERABLE

i. *Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e., adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?*

Yes. See previous sections.

ii. *What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?*

Section 4 and table 3 show a clear example of the applied methodology for using standard vocs and making data interoperable.

iii. *Will you be using standard vocabularies for all data types present in your dataset, to allow inter-disciplinary interoperability?*

Yes. See previous sections.

iv. *In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?*

The adopted NVS vocabulary service offers both NVS RESTful, SOAP and SPARQL services. AGitHub repository for key NVS vocabularies[18] tracks the discussion on new terms adoption. NAUTILOS partners are already using these services and are collaborating with SeaDataNet and linked projects (e.g., EMODnet).

The NVS service is also open to map and manage new terms and ontologies, therefore the primary approach of NAUTILOS will be to interact with the service (and the people managing the vocabulary) to have a community definition, acceptance and hence adoption of new proposed terms.

[18] github.com/nvs-vocabs

4. INCREASE DATA RE-USE (THROUGH CLARIFYING LICENSES)

i. How will the data be licensed to permit the widest re-use possible?

CC BY-SA or CC BY-NC licenses. In some cases, to be identified, and depending on the data/document, CC BY or CC0 (public domain) can be taken into consideration.

ii. When will the data be made available for re-use? 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.

The general approach is presented in the previous sections. In general, data will be made available as soon as possible, with the latest being the end of the project.

iii. Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

Data and products produced or used in the project are encouraged to be used and adopted by third parties as soon as possible. We are certain that all data produced in this project will be of importance to the wider scientific community and to future projects.

iv. How long is it intended that the data remains re-usable? Are data quality assurance processes described?

Data quality and data quality flag approaches are documented. Moreover, versioning/tagging of the data and products is applied to facilitate long term reusability.

Validated NAUTILOS data will be assigned with a permanent digital object identifier ("doi") - according to DataCite[19] model. Having a data-doi enables data provenience traceability, citation and acknowledgment, as well as long term availability and reusability of data. DataCite repositories such SEANOE, PANGAEA, Zenodo (and many others) have the commitment to maintain the data repositories.

As described in the previous sections, NAUTILOS DMP describe the methods to organize the data during the project life (see also D8.3) and how to provide and link NAUTILOS data towards European and Global integrators where the entry point for this interoperability action (i.e. EMODnet Ingestion) is designed to offer data with doi whenever data has not it yet. According to the NAUTILOS DMP, NAUTILOS data has to be accompanied by metadata and metadata reports on the applied data quality procedure. These metadata do not describe the procedures themselves, but they regulate and provide a field with a link to the applied procedure . These simple procedures are the key cornerstones for ensuring project' data reusability and data long term legacy and traceability.

[19] https://datacite.org/value.html

v. Further to the FAIR principles, DMPs should also address:

In the European legislative context, NAUTILOS encourages data providers to comply with metadata encoding following the requirements of the INSPIRE directive as well as the Directive 2003/4/EC[20]. Therefore, INSPIRE compliant XML formats such as ISO 19115 or accordingly enriched simpler formats such as extended Dublin Core are the preferred metadata profile. If metadata is integrated within a data file, the file needs to be in an agreed format such as OceanSites NetCDF format. The use of ERDDAP is combining and solving this need.

[20] DIRECTIVE 2003/4/EC „on public access to environmental information and repealing Council Directive 90/313/EEC"

V. ALLOCATION OF RESOURCES

i. What are the costs for making data FAIR in your project?

Besides the development and hosting of the data infrastructure that enables the compliance with FAIR principles, mainly allocated by the partner ETT, all the partners have allocated a dedicated budget for the promotion and dissemination activities.

ii. How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

Wherever the project will be presented, the presentations will report links to the NAUTILOS data infrastructure and instructions and information on how NAUTILOS implements the data and products Findability, Accessibility, Interoperability and Reusability.

iii. Who will be responsible for data management in your project?

Data Management is included in a dedicated project work package (WP8) and responsible for the data management is Antonio Novellino, who has long experience in data management and ocean data management and sharing. Following is his expertise history in brief.

He holds a PhD in Biotechnology and Bioengineering, MSc Biomedical Engineering, Certified Data Scientists. From 2008 to 2010, he served the European Commission, JRC – IHCP as a senior researcher. Manager in R&D/Smart Sustainable Cities at ETT. He serves on the scientific board of the Ligurian Cluster of Marine Technology DLTM (www.dltm.it) and the board of Consortium TRAIN. He is member of the EuroGOOS DATAMEQ group and contributes to several EuroGOOS Task Teams for advising (http://eurogoos.eu/) on operational oceanography data management procedures and standards. He is member of the ONTM (National Observatory for Marine Technologies) Innovation Hub. He is member of the SOOS DMSC (South Ocean Observing System Data Management Steering Committee) and DOOS DMTT (Deep Ocean Observing System Data Management Technical Team). He serves on the EMODnet Steering Committee, the EMODnet Technical Working Group. He serves the Expert Team on WIS Centres (ET-WISC) and Task Team on Data Centres (TT-DC). He is the EMODnet Physics coordinator and Copernicus Marine Service DU deputy coordinator. He is proactively involved major EU projects in the field of oceanography and marine data management (EMODnet Ingestion, EMODnet Chemistry, AtlantOS on H2020 BG8, SeaDataCloud on H2020, EuroSEA on H2020, Jerico S3 on H2020, SO-CHIC on H2020, NAUTILOS on H2020, BlueCloud2026 on HEurope, OCEANICE on HEurope, OLAMUR on HEurope Mission Restore, EFFECTIVE on HEurope Mission Restore). He was lately appointed as EOSC-FAIR-Champion

iv. Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

For the duration of the project, data is going to be hosted by the NAUTILOS data management infrastructure. This infrastructure is going to be interoperable with the key European integrating data ocean infrastructures and programs (CMEMS, EMODnet and SeaDataCloud) and new data will directly be harvested by these long-term safe keeping systems. It should be noted that EMODnet and in particular by means of the EMODnet Data Ingestion and Safe Keeping project, is implementing a dedicated action to collect and provide long-term preservation of ocean data and NAUTILOS DMP already consider this system for its data legacy (see also section 4.iv). Moreover, NAUTILOS designed a specific Task (Task 9.5) to further investigate and describe a procedure to deal with data legacy and integration into European platforms. As already stated, a specific goal of NAUTILOS project and DMP is to perform long term data stewardship and data legacy in collaboration with other key European networks such as the SeaDataNet network of National Oceanographic Data Centres and IODE centres.

VI. DATA SECURITY

i. What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

The NAUTILOS infrastructure is going to be deployed on the ARUBA.it infrastructure[21]. Since 2015 Aruba is running a dedicated service to private business clients and it provides the client with top level services such as Data Centre (Virtual Servers, Real Servers, hosting infrastructures), Back up and Disaster Recovery etc. ARUBA also provides us with most recent services for data security cryptography (AES), security protocols (AES, SSL) and bandwidth balance. The main characteristics of the service are:

SLA	99,80%
security	crypted transmission channel
	(optional) storage crypting AES-256
min backup timing	1h
schedule	anytime
	granularity – single backup job
Backup Account number	unlimited
concurrent agents	depends on the agreed service
max number of backup jobs	unlimited
bandwidth	unlimited (upload/download)
Certifications	ISO 9001:2015, ISO 27001:2013
service desk	24h
Cloud security certification	ISO/IEC 27017:2015
data privacy	ISO/IEC 2018:2014
security incidents:	ISO/IEC 27035:2016

ii. Is the data safely stored in certified repositories for long-term preservation and curation?

While the ARUBA infrastructure is going to guarantee the data security, back up and disaster recovery services, for long term preservation and curation, as described in the previous paragraph, the NAUTILOS project will work in collaboration with the major ocean data infrastructure, which has the mandate for long-term data curation and preservation: EMODnet and in particular the Data Ingestion and Safe Keeping project, ICES, SeaDataNet networks of NODCs, CMEMS. The NAUTILOS partners are also involved in other global initiatives (SOCAT, GLODAP, etc) and NAUTILOS data will be also shared (and clearly NAUTILOS labelled) and made available towards these infrastructures.

[21] https://business.aruba.it/azienda.aspx

VII. ETHICAL ASPECTS

i. Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

All the ethical aspects of this project are covered under WP13 and its deliverables.

ii. Projects participating to the ORDP might present information relevant to the ethical aspects (data protection) in the DMP. In such a case, the ethics chapter of the DoA may simply refer to the DMP for more information on the details of the ethics aspects related to data.

All the ethical aspects of this project are covered under WP13 and its deliverables.

iii. Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Whenever the NAUTILOS project will implement surveys, questionnaires, or collect personal data for any reason (e.g., attendance to organized events), European GRDP law will be used as reference and the user will be informed about the use of personal data. In general, NAUTILOS will not transfer personal data (e.g., email addresses) to other entities and the only use will be setting up a distribution list to inform users about project progress. User will be always able to change his consensus and ask for being removed from the distribution channel. More details, as well as the templates prepared for the Ethics Work Package are available as the deliverable documents for WP13, in particular D13.1 and D13.2 dealing with personal data and procedures and criteria that will be used to identify/recruit research participants, as well as the informed consent procedures that will be implemented for the participation of humans external to NAUTILOS to the project activities and in regard to their data processing. As previously mentioned and following the establishment of the Ethical Advisory Board in NAUUTILOS, an update of the informed consent forms has been prepared and submitted as annexes to deliverable D13.7. The updated forms are available in different languages according to the needs of the various NAUTILOS partners, who will have to carry out activities related to personal data.

VIII. OTHER ISSUES

i. Do you make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones?

Nothing to report.

ii. Further support in developing your DMP

Nothing to report.

IX. APPENDIX 1: REFERENCES AND RELATED DOCUMENTS

ID	Reference or Related Document	Source or Link/Location
1	GDPR	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679
2	EU Regulation	https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e3722-1-1
3	ISO 27001	*https://en.wikipedia.org/wiki/ISO/IEC_27001*
4	ENISA - Handbook on Security of Personal Data Processing - 2017	ISBN 978-92-9204-251-6, DOI 10.2824/569768
5	European Data Protection Board (2020).	https://edpb.europa.eu/sites/edpb/files/consultation/edpb_recommendations_202001_supplementarymeasurestransferstools_en.pdf
6	European University Institute (2019). Guide on Good Data Protection Practice in Research	https://www.eui.eu/documents/servicesadmin/deanofstudies/researchethics/guide-data-protection-research.pdf