www.nautilos-h2020.eu

Organisation CNR
Department ISTI

Data Management Plan

Date: 09.03.2021
Doc. Version: V3

Document Control Information

Settings	Value
Deliverable Title	Data Management Plan
Work Package Title	Project Management
Deliverable number	D1.3
Description	In this deliverable the Data Management Plan (DMP) of the project will be written, in compliance with the H2020 Data Management Guidelines, also based on inputs from WP8. It will outline a data management policy, including data to be generated by the project, their potential exploitation, curation and storage. Additionally, in line with the principles of Open Access to research data and publications generated through H2020 programmes, NAUTILOS will participate in the Open Research Data Pilot carried out by the European Commission.
Lead Beneficiary	CNR
Lead Authors	Antonio Novellino
Contributors	Federica Colombo, Anna Gianvincenzo, Gabriele Pieri, Marco Tampucci
Submitted by	Gabriele Pieri
Doc. Version (Revision number)	V3
Sensitivity (Security):	Public
Date:	09/03/2021

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	Approver	*Approved after revision*	08/03/2021

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	18/01/21	Antonio Novellino	First complete draft
V2	05/03/21	Gabriele Pieri	Re-editing and additions
V3	09/03/21	Gabriele Pieri	Minor details fixed 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: http://www.nautilos-project.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. For this reason, the NAUTILOS data management policy is committed to make available NAUTILOS data via both the project website and major initiatives, and data portals such EMODnet, CMEMS, etc (where NAUTILOS partners are involved). 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 to data and metadata produced by their facilities and are committed to working towards the implementation of this principle;
- Data and metadata generated during the project will be made available following an open access policy, without any restrictions and available for free to third parties;
- Appropriate controlled dictionaries (CF convention and SeaDataNet vocabularies) are recommended to be used for metadata description;
- A metadata catalogue will be accessible on the NAUTILOS portal and data will be provided to ongoing projects and initiatives such as EMODnet, CMEMS, SeaDataNet, etc.

Contribution of data:

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

- 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;
- Data providers are requested to inform of any national policies that may place special conditions on the redistribution of data;
- 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 available within the data integrator portals and initiatives (EMODnet, CMEMS, etc).
The NAUTILOS partners are not responsible for any use and misuse made by end-users.

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

Abbreviation	Definition
BODC	British Oceanographic Data Centre
CDI	Common Data Index
CF convention	Climate and Forecast convention
CMEMS	Copernicus Marine Environment Monitoring System
DAC	Data Archive Centre
DMP	Data management policy
EDMO	European Directory of Marine Organisations
EMODnet	European Marine Observation and Data Network
GOOS	Global Ocean Observing System
GTS	Global Telecommunication System
ICES	International Council for the Exploration of the Sea
INSPIRE	Infrastructure for Spatial Information in Europe
JERICO	Joint European Research Infrastructure of Coastal Observatories
MEDIN	Marine Environmental Data and Information Network
NetCDF	Network Common Data Form
OGC	Open Geospatial Consortium
QC	Quality control
SOOS	Southern Ocean Observing System
TAC	Thematic Assembly Centre
WMO	World Maritime Organisation

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.

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 will follow 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 Data Archive Centres (DACs) and data management infrastructures (e.g., EMODnet) in their respective accepted formats.

Data management will be developed in coordination and collaboration with already existing infrastructures and integrators (e.g., EMODnet, CMEMS, SOOS) to avoid effort duplication and to facilitate a fast adoption and availability of the produced data.

In this framework, NAUTILOS data management policy (DMP) includes and clarifies the roles on the ownership and custodianship of the data, as well as the recommendations for data

citation. NAUTILOS partners are not responsible for any use and misuse made by end-users. Following the structure of the Horizon 2020 DMP template[2], this document presents the first version of the NAUTILOS DMP, it defines the general policy and approach to manage data and handles data management related issues on the administrative and technical level. In this framework, the NAUTILOS DMP will evolve during the lifespan of the project and next versions will present more details.

III. DATA SUMMARY

As the data identification and collection activities are still ongoing, the initial DMP can currently only provide an incomplete picture of the datasets that are needed. Nevertheless, while the focus of the first version of the DMP is mainly on existing data, the next 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.
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;

[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.

- 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?

This 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.

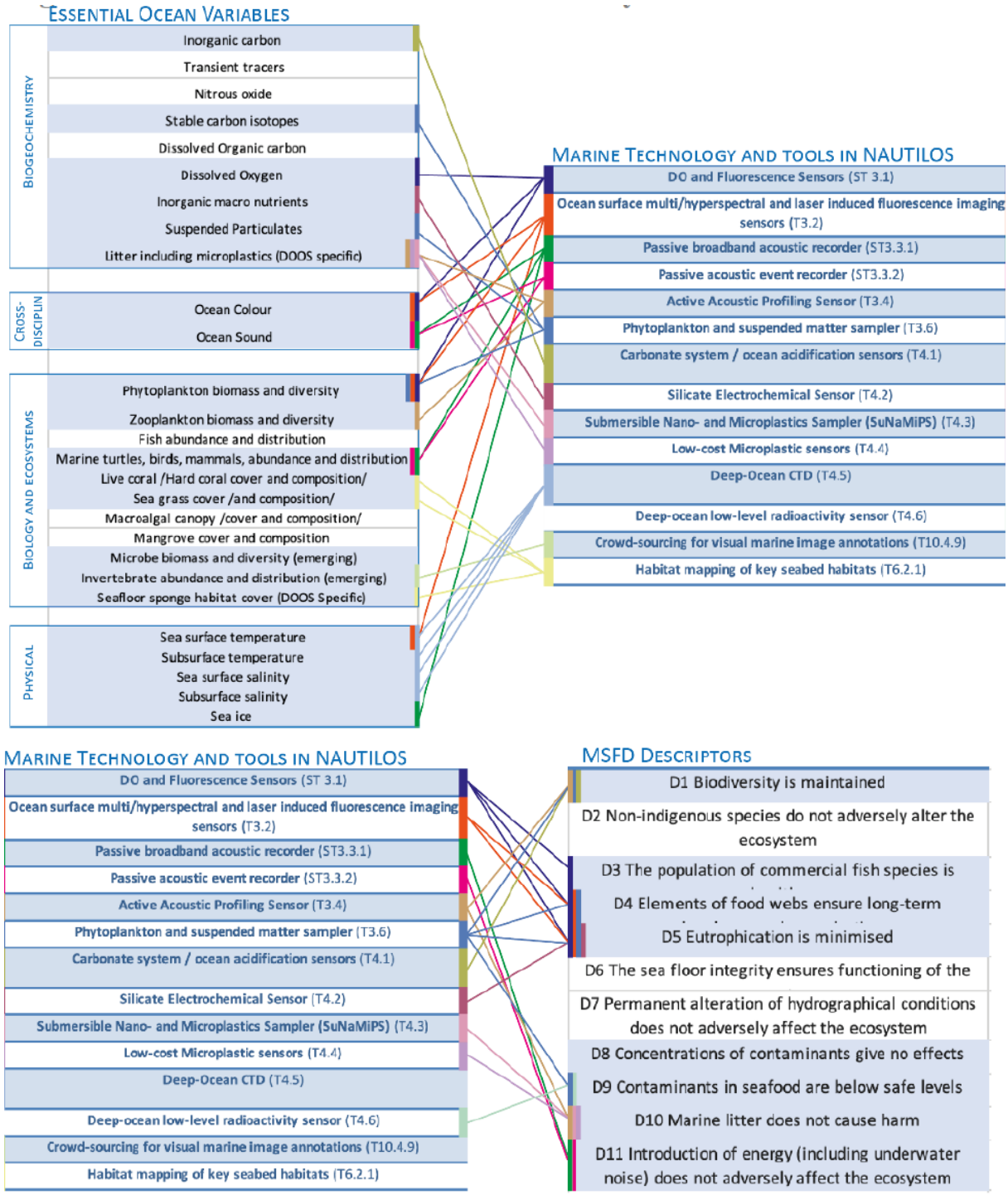

Figure 1 Environmental variables covered by NAUTILOS in terms of: Essential Ocean Variables (top); and MSFD descriptors (bottom) covered.

There will be 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 will consist 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 quality 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, do not have a well-defined data stream yet. NAUTILOS will operationally implement the data management and data transfer to the appropriate DAC or identified official European data repositories.

For the well-established parameters, NAUTILOS data management will be developed in coordination and collaboration with already existing infrastructures (DACs and Global DACs) and integrators (e.g., EMODnet, CMEMS, JERICO[4], ICES[5], OGC[6], DarwinCore[7], etc). This approach will avoid effort duplication and will facilitate a fast and smooth integration into European marine data infrastructures. The NAUTILOS project will work on the relevant standards from the outset (INSPIRE[8], OGC), and promptly supply relevant field data to the Member States (e.g., BODC[9], MEDIN[10]), European (e.g., CMEMS), and International bodies (e.g., GOOS[11]). The resulting environmental data will be processed to a form that meets widely accepted data standards such as MEDIN discovery metadata standard, INSPIRE data specification, etc. For new parameters or data with a less structured data flows, the project will design and develop specific Thematic Assembly Centres (TACs) by adopting and adapting best practices developed for other parameters and anticipating the European integrators and infrastructures.

Concerning 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	This document. Public ORDP: Open Research Data Pilot document.

[3] https://eurogoos.eu/data-management-exchange-quality-working-group-data-meq/
[4] https://www.jerico-ri.eu/
[5] https://www.ices.dk/Pages/default.aspx
[6] https://www.ogc.org/
[7] https://dwc.tdwg.org/
[8] https://inspire.ec.europa.eu/
[9] https://www.bodc.ac.uk/
[10] https://www.medin.org.uk/
[11] https://www.goosocean.org/

D1.10 Data Management Plan – 1st periodic report update	Update of DMP at M18. Public ORDP: Open Research Data Pilot document.
D1.11 Data Management Plan – 2nd periodic report update	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 due at the end of 1st year.
D8.4 Design of Thematic Assembly Center for innovative parameters	Description of how to set up a dedicated NAUTILOS assembly centre for new parameters. Other: design of software platform due at the end of 1st year.
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 due 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 due 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[12]. 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.). For the historical validated datasets (fixed stations – mooring, tide gauge), the metadata format is the CDI (common data index) and the transport formats are ODV4 and NetCDF (CF convention - Climate and Forecast Metadata Convention).

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:

[12] http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf

- 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[13] are recommended;
- Platform types from SDN L06;
- Intuition codes: EDMO (European Directory of Marine Organisations)[14].

2.2. 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.

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

NAUTILOS will use relevant publicly available oceanographic dataset, such as those made available by key European data integration infrastructures like CMEMS INSTAC, EMODnet Physics, SeaDataNet.

[13] https://vocab.seadatanet.org/v_bodc_vocab_v2/vocab_relations.asp?lib=P02

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

The analysis is going to be extended to international data networks (e.g., ARGO[15], OceanSITES, etc.). Another source of data will be satellite remote sensing from both European and international space agencies. Furthermore, any additional data sources that will be identified and which are not yet included in any of the cited infrastructures will be added in the project Data Inventory.

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.

Data is originated by the 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).

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.

[15] http://www.argo.net/

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[16], 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 identification of the sensors, their configuration, the deployment at sea and collection of the observation that are regularly transmitted to a shore-located receiving station. The receiving station reformats the data messages, applying some automated quality control, and then distributes the data over one of the several communication pathways, like the Global Telecommunication System (GTS). Data Assembly Centres collect these data and develop added value reanalysis/products. Each step produces information that may be

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

associated to the given observation, hence generating an immense volume of metadata information.

Data management associated with the ocean observing systems is hence scattered and requires a massive coordination effort. In this fragmented landscape, interoperability is essential to enable data to be used, reused, stored, and integrated. The building block for an integrated system in which components can exchange and understand information is the standardisation of formats, distribution protocols and metadata. In this framework, NAUTILOS will be acting as the integrator by connecting different data sources. It will apply a minimum number of metadata elements designed to provide users with information identifying a collection of files as a thematic/coherent dataset. It will support the search through that collection 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 Standard[17] 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.

Some examples are Climate and Forecasting conventions for parameters that are mapped versus the NVS P01-P02, and the adoption of CMEMS INSTAC convention and recommendations for unique platform identification. Each platform in the CMEMS INSTAC has a unique code that, when available, is the WMO (World Maritime Organisation) code. OCEANOBS (as the focal point for the practical coordination of the in-situ ocean observing systems defined by JCOMM) also has a key role in the overall management of the platforms and programs metadata, NAUTILOS will interact with OCENAOBS and adopt and follow international recommendations.

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

Following GOOS recommendation 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 will adopt ERDDAP[18] as the core solution for data management. The goal is to make easier for users to access scientific data.

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

ERDDAP implements FGDC Web Accessible Folder (WAF) with FGDC-STD-001-1998 and ISO 19115 WAF with ISO 19115-2/19139. The second core data interoperability technology in NAUTILOS is 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.

The main advantages and benefits of using ERDDAP are:

- ERDDAP acts as a middleware between the end user and various remote data servers. When the user requests data from ERDDAP, ERDDAP reformats the request into the format required by the remote server, sends the request to the remote server, gets the data, reformats the data into the format requested, and sends the data to the user. The user no longer has to go to different data servers to get data from different datasets.
- ERDDAP offers an easy-to-use, consistent way to request data via the OPeNDAP standard. Many datasets can also be accessed via ERDDAP's Web Map Service (WMS).
- ERDDAP returns data into the common file format of choice. ERDDAP offers all data as .html table, ESRI .asc and .csv, Google Earth .kml, OPeNDAP binary, .mat, .nc, ODV .txt, .csv, .tsv, .json, and .xhtml.
- ERDDAP can also return a .png or .pdf image with a customised graph or map.
- ERDDAP standardises the dates+times in the results. Data from other servers are hard to compare because the dates+times often are expressed in different formats (for example, "Jan 2, 2018", 02-JAN-2018, 1/2/18, 2/1/18, 2018-01-02, "days since Jan 1, 1900"). For string times, ERDDAP always uses the ISO 8601:2004(E) standard format, for example, 2018-01-02T00:00:00Z. For numeric times, ERDDAP always uses "seconds since 1970-01-01T00:00:00Z". ERDDAP always uses the Zulu (UTC, GMT)

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

time zone to remove the difficulties related to working with different time zones and standard time versus daylight saving time.

- ERDDAP has web pages (for humans with browsers) and RESTful web (for computer programs). You can bypass ERDDAP's web pages and use ERDDAP's RESTful web services (for example, for searching for datasets, for downloading data, for making maps) directly from any computer program (for example, MATLAB, R, or a program that you write) and even from web pages (via HTML image tags or JavaScript).
- ERDDAP is free and open-source code (JAVA program and source code is available in GitHub[19]) that uses Apache compatible software licenses.

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 follow a NAUTILOS naming convention where the file name already provides the user with information of contents.

The following tags will be used:

SOURCE.PROVIDER_

DISSEMINATION.CHANNEL_

PROCESSING.LEVEL_

PARAMETER_

PLATFORM_

PLATFORMTYPE_

DATATYPE_

TIMEFREQUENCY.TIMERESOLUTION_

e.g., NAUTILOS_ERD_L2_TEMP_GLIDERXXX_GL_PR_30m_1day

meaning that the source is NAUTILOS project, it is distributed by the NAUTILOS ERRDAP, it is a L2 processing level (as defined for remote sensing), it is a Temperature Profile dataset recorded by GliderXXX that is a glider, with a sample every 30m for 1 day.

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.

[19] https://github.com/BobSimons

iv. Do you provide clear version numbers?

All NAUTILOS products will be clearly labelled.

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 will follow 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[20].

The applicable recommendations[21] relying on existing international standards that will ensure cross platform coherence and facilitate data discovery for users and data integration:
"*Platforms should have a unique identifier that will be either WMO for most platforms or ICES code for ships (and links to expocodes)*"

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-P09 (parameter), P07 (CF variable), P06 (units) L06 (platform types) from SeaDataNet controlled vocabularies managed by NERC/BODC (Vocabulary Server (version 2.0).
Moreover, within the AtlantOS project An AtlantOS Essential Variables list of terms (aggregated level), related to ECV –EOV or other, has been defined and was published in June 2016 on the NERC/BODC Vocabulary Server (version 2.0) as A05 vocabulary (https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/).

Institutions listed in a data file shall be identified by a unique code from the EDMO existing catalogue. Quality control (QC) information will be attached to the data; both quality flags that can be mapped to SeaDataNet flag scale (available in the SeaDataNet Common Vocabularies (http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx/) as list L201), and, whenever known, processing level information ("qualified in RT using automated procedures" or "processed in delayed DM by Scientist").
Below is an example of an extract of the mapping defined for CMEMS INS TAC/EMODnet Physics between parameters in the data format and (P01 parameters + P06 units) in BODC controlled vocabularies.

Table 2. Example of metadata mapping

param	long_name	cf_standard_nam	unit	sdn_parameter_u	sdn_parameter_	sdn_uom_u	sdn_uom_uri

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

[21] https://archimer.ifremer.fr/doc/00360/47090/47013.pdf

		e		r	ur	rn	
PRES	Sea water pressure, equals 0 at sea-level	sea_water_pressure	decibar	SDN:P01::PRESPR01	http://vocab.nerc.ac.uk/collection/P01/current/PRESPR01/	SDN:P061::UPDB	http://vocab.nerc.ac.uk/collection/P06/current/UPDB/
TEMP	Sea temperature in- situ ITS-90 scale	sea_water_temperature	degree_Celsius	SDN:P01::TEMPST01	http://vocab.nerc.ac.uk/collection/P01/current/TEMPST01/	SDN:P061::UPAA	http://vocab.nerc.ac.uk/collection/P06/current/UPAA/

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.

Part of the data and data products will have restricted access during a preliminary phase of the project. During this phase, these data and data products will be freely accessible to the consortium partners only. This private phase is necessary to make sure data are released for public access only after the required processing and quality control have been performed, as well as allowing partners some time to report results in scientific journals. NAUTILOS will implement a data management infrastructure composed of two mirror systems in which the first one will allow access only to project partners and the second one that is totally open to any user. As soon as the internal validation and assessment procedures are closed, the data/products will be moved from the first system to the second.

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 will implement the most recent dissemination catalogues and technologies and will link and deliver data into the most relevant ocean data management infrastructures and programs.

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

Data will be open and freely available and can be viewed and used by using well-known software tools.

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 eventual 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, other general purpose but well-established repositories (e.g., Zenodo) will be considered.

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?

Before the full disclosure of the data, only project partners can access data and product. Access will be managed by an ad hoc authentication system.

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 will be published, they will be public and accessible without any restriction nor authentication needed. NAUTILOS will monitor the typology of user (in compliance with GRDP) by means of a web form that would be filled on voluntary base. 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.

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?

A preliminary list of the vocabularies to be adopted in NAUTILOS is in Table 2

Table 3. Metadata Vocabularies

Metadata field	**Vocabulary exists**	**Link to vocabulary**	**Vocabulary governance**
Platform	Yes	http://vocab.nerc.ac.uk/collection/L06/current/	BODC
Platform short diagram	yes	CMEMS INSTAC – EMODnet Physics	EuroGOOS
contributors_role			NAUTILOS
naming_authority	Yes	https://edmo.seadatanet.org/	SeaDataNet
Institution	Yes	https://edmo.seadatanet.org/	SeaDataNet
rtqc_method	No		NAUTILOS
platform_type	Yes		INSTAC
ICES_code	Yes	https://ocean.ices.dk/codes/ShipCodes.aspx	ICES
sensor_model	Yes	http://vocab.nerc.ac.uk/collection/L22/current/	BODC

data_mode	No		NAUTILOS
Phase	No		NAUTILOS
variable names	Yes	http://vocab.nerc.ac.uk/collection/P02/current/	BODC
Time	yes	ISO8601	ISO
Datum	Yes	WGS84	ISO

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?

Controlled vocabularies consist of lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. Using standardised sets of terms solves the problem of ambiguities associated with data mark-up and enables records to be interpreted by computers. This opens datasets to a whole world of possibilities for computer aided manipulation, distribution and long-term reuse. In this framework, an important outcome of a past ICES/IOC Study Group on the Development of Marine Data Exchange Systems was the establishment of a combined SeaDataNet and MarineXML Vocabulary Content Governance Group (SeaVoX), moderated and maintained by NERC-BODC (NERC Vocabulary Server – NVS). SeaVoX operates by a mailing list server, open to anyone with a genuine interest in controlled vocabularies for the marine science domain (it has an active membership of experts from SeaDataNet, IOC/IODE, ICES and other international groups).

The NVS offers both NVS RESTful, SOAP and SPARQL services. AGitHub repository for key NVS vocabularies[22] 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.

[22] 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, we intend to apply versioning of the data and products to facilitate the re-usability of the data for the longest period possible. Data will be stored in data repositories and in principle will be available and re-usable indefinitely.

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[23]. 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.

V. ALLOCATION OF RESOURCES

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

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 employment history in brief.

He holds a PhD in Biotechnology and Bioengineering, MSc Biomedical Engineering, Research Manager at ETT, Coordinator of EMODnet Physics. He has been involved in R&D projects at both National and European level for the past 15 years. From 2008 to 2010, he served the European Commission, JRC – IHCP as a senior researcher. At present, he is the ETT Research Manager and he is coordinating the R&D department activities. He serves on the Board of Directors of Consortium Si4Life (manager of Polo Si4Life - Life Sciences Liguria Region - www.si4life.com), he served on the board of Consortium Tecnomar (about 100 companies working on maritime and environment technology, www.consorziotecnomar.com). He serves the techno-scientific board of the Ligurian Cluster of Marine Technology DLTM (www.dltm.it) and the board of Consortium TRAIN (innovation in energy and transport management, www.consorziotrain.org) and is member of the EuroGOOS DATAMEQ group for advising (http://eurogoos.eu/) on operational oceanography data management procedures. He serves 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 co-author of scientific publications and he participates as reviewer for scientific journals for Elsevier and Frontiers. He proactively supported EGU ESSI 1.1/OS4.35 session since 2014. He is the EMODnet Physics coordinator and CMEMS DU deputy coordinator.

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 linked to the key European integrating data

ocean infrastructures and programs (CMEMS, EMODnet and SeaDataCloud) and new data will directly flow into 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. A specific Task (Task 9.5) to deal with data legacy and integration into European platforms is foreseen amongst NAUTILOS activities. A specific goal will be 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[24]. 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

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

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.

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.

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