Organisation ETT SpA

D8.1

Technical documentation and operational field primary data capture system

Date: 08/04/2021
Doc. Version: 2.0

10.5281/zenodo.7211768

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

Document Control Information

Settings	Value
Deliverable Title	Technical documentation and operational field primary data capture system
Work Package Title	Data Management
FDeliverable number	D8.1
Description	The Deliverable collects a review of standards and best practices to be applied to field operation and data collection, e.g. sensor calibration, data cross-platform validation etc. It takes into consideration the guideline from European Networks (EuroGOOS WGs/TTs, EMODnet, SeaDataNet, CMEMS, AtlantOS/EuroSEA, JERICO NEXT/S3) as well as international ones (OceanGliders, ARGO, AniBOS, etc.) and provides the reader links to find documentation and best practices to be applied.
Lead Beneficiary	ETT SpA
Lead Authors	A. Novellino, F. Misurale
Contributors	
Submitted by	G. Pieri
Doc. Version (Revision number)	[2.0]
Sensitivity (Security):	Public
Date:	[08/04/2022]
doi	10.5281/zenodo.7211768

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
Flávio Martins	Reviewer	Approve	07/04/2022

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
1.0	01/03/2022	F. Misurale	Preliminary draft
1.1	15/03/2022	A. Novellino	General Revision
1.2	21/03/2022	F. Misurale	General Revision
2.0	04/04/2022	A. Novellino	Final version

Configuration Management: Document Location

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

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

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

Acknowledgement

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

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

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

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

More information on the project can be found at: https://www.nautilos-h2020.eu/

Copyright

TABLE OF CONTENTS

Executive Summary

The large variety of platforms and sensors globally deployed and available, allows access to a conspicuous quantity of data and variables with heterogeneous formats and scales.

In the last decades governments and the scientific world have encouraged standardizations and best practices for data collection, elaboration and distribution. Several initiatives had led the bases for globally shared practices and quality indexes.

This deliverable aims at giving a state of the art concerning the existing best practices and quality indices of the most important platforms, initiatives and projects for the scope of NAUTULOS. The present document is located in the context of WP8 and intended to be in agreement with other deliverables concerning exploitation strategy (D11.1), data management (D8.3), interoperability (D8.5, D8.2 and D2.3), quality (D1.4) and techniques (D2.2 and D8.4).

The final part of the document is dedicated to recommendation for document sharing between project partners and outside the project.

LIST OF FIGURES

LIST OF TABLES

List of Acronyms and Abbreviations

Abbreviation	Definition
ANIBOS	Animal-Borne Ocean Sensors
ASV	Autonomous Surface Vehicles
AUV	Autonomous Underwater Vehicles
BGC	Bio-Geo-Chemical
CMS	Copernicus Marine Services
DAC	Data Assembly Center
DBCP	Data Buoy Cooperation Panel
DMP	Data Management Plan
DOI	Digital Object Identifier
DOOS	Deep-Ocean Observing Strategy
EC	European Commission
EMODNET	European Marine Observation and Data Network
EOOS	European Ocean Observing System
EOV	Essential Ocean Variables
Euro-Argo	European contribution to the international Argo Program
EuroGOOS	European Global Ocean Observing System
EuroSea	Improving and integrating the European Ocean Observing and Forecasting System
FAIR	Findability, Accessibility, Interoperability, and Reusability
FOO	Framework for Ocean Observing
GLOSS	Global Sea Level Observing System
GOOS	Global Ocean Observing System
GOSHIP	Global Ocean Ship-Based Hydrographic Investigations Programme
HELCOM	Helsinki Commission (Baltic Marine Environment Protection Commission)
GRA	GOOS Regional Alliances
ICES	International Council for the Exploration of the Sea
INSPIRE	INfrastructure for SPatial InfoRmation in Europe
IOC	Intergovernmental Oceanographic Commission
IODE	International Organization for Data Exchange
KPI	Key Performance Indicator
MEOP	Marine Mammals Exploring the Oceans Pole to Pole
OBPS	Ocean Best Practices System
RI	Research Infrastructure
SOT	Ship Observations Team

1. INTRODUCTION

The value of any system is contingent on its relationship to its user community and scaled by the degree of its adoption.

To this end, during the last decades the International Organization for Data Exchange (IODE), the Framework for Ocean Observing (FOO), the Global Ocean Observing System (GOOS), as well as European initiatives (Copernicus Marine Service, European Marine Observations and Data network, SeaDataNet networks of National Oceanographic Data Centers), etc. have encouraged standardizations and best practices for data collection, elaboration and distribution.

More recently, the Ocean Best Practices System (OBPS) supports the objective of fostering and making available best practices across the value chain, from observations to data management and applications for understanding the state and potential of our ocean. To gain maximum value from this vital information we need to use commonly accepted methods, 'best practices', across our ocean observing system, in order to deliver data that is interoperable, re-usable, of high quality and with consistent latency.

The OBPS workflow put the users at the center of the process. They may be able to access best practice from running networks, training to learn about best practices, have the possibility to develop, publish and let other users to adopt best practices.

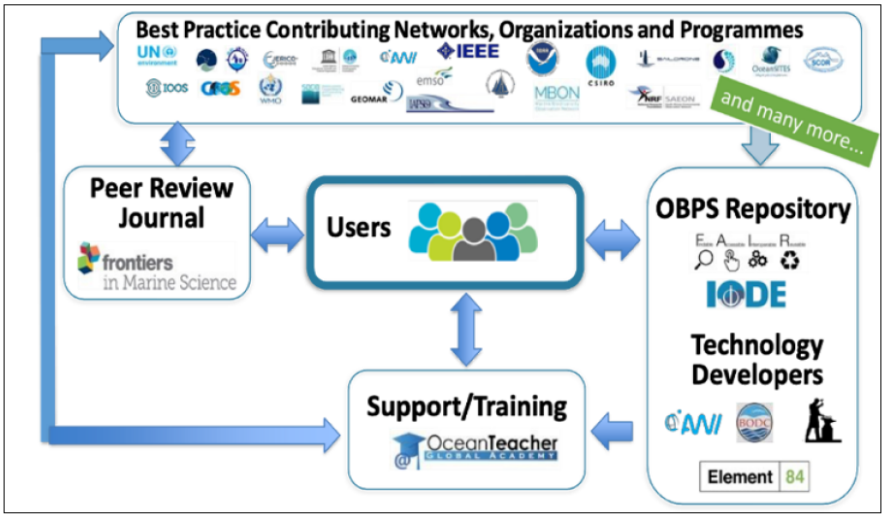

Figure 1. OBPS diagram

NAUTILOS has the strategic objective of filling 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 is 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.

This deliverable recalls the technologies and sensors that NAUTILOS is developing, then lists the running international initiatives also included in the OBPS, describes the approach to facilitate NAUTILOS data flow into European Marine depositories and initiatives.
This activity is also supporting WP1 (DMP) and T8.2 as one main NAUTILOS goal is to enable and facilitate project data use and reuse.

Last section presents how to apply best practice in internal document management. The procedure applies to documentation - all documented outputs of the project have to be captured in an open access format and made immediately available to all project partners – and to data and data products. The project is moving towards its sensors' deployment and data acquisition phase hence the production of valuable environmental data.

The management of internal documents is organized around two main tools, a project Google Team Drive – which provides the partners an easy tool to co-work on documents in parallel and from anywhere – and a project OwnCloud service that provides the project with a repository of the official project documentation.

2. NAUTILOS TECHNOLOGIES AND REFERENCE NETWORKS

NAUTILOS is focusing on 17 instrumentation/tools that can operate from shallow coastal waters to open and deep-sea sites, providing complete datasets for studying the marine ecosystem functions and advanced data products and tools.

NAUTILOS is developing sensors to match and fit needs of operational oceanography platforms, more specifically it is targeting the following integrating platforms:

- Fisheries Observing Systems (Opportunity ships);
- Aquaculture Observing Systems (Ferrybox)
- Acoustic Marine Mammal Monitoring System (fixed point observation);
- Autonomous vehicles and Lander missions (Gliders/AUV)
- ARGO floats
- Animal- borne instrumentation
- Mooring buoys and fixed observatories

These platforms already fall into or can be associated to one of the GOOS platforms networks (Data Buoy Cooperation Panel (DBCP)[1], Global Ocean Ship-Based Hydrographic Investigations Programme (GO-SHIP)[2], Ship Observations Team (SOT)[3], ARGO[4], OceanSITES[5], OceanGliders[6], Animal-Borne Ocean Sensors (AniBOS)) that organize the standards and best practices for the platform stakeholders.

At European level, EuroGOOS that has established Task Teams, and TT members collaborate in the areas of shared priorities, exchange best practices, and feed data to the EuroGOOS ROOS regional portals, and European Marine data programs and initiatives: EMODnet, and Copernicus Marine Service and SeaDataNet.

The European Marine Data programs and initiatives largely contribute to standards and best practices definition and in this framework most of the parameters and platforms have a precise and well consolidated policy of quality and data/metadata codification and format.

Some of these initiatives also deal with the long-term preservation and legacy of (some) data and in particular the IODE network of National Oceanographic Data Centers, and well-established international repositories (ICES DB, PANGAEA, …) were designed and are evolving to support researchers with these crucial topics of the ocean data research activities.

[1] https://www.ocean-ops.org/dbcp/
[2] https://www.go-ship.org/
[3] https://www.ocean-ops.org/sot/
[4] https://argo.ucsd.edu/
[5] http://www.oceansites.org/
[6] https://www.oceangliders.org/

To maximize the value and impact data collection, NAUTILOS strategy has to consider these end-points, the retrieval process for data and/or metadata production, and to apply these standards/requirements backwards toward the data dissemination/production.

NAUTILOS is covering 14 Biology and Ecosystem and Biogeochemical EOVs (inorganic carbon, stable carbon isotopes, dissolved oxygen, inorganic macronutrients, suspended particulates, ocean colour, ocean sound, phytoplankton biomass and diversity, zooplankton biomass and diversity, marine turtles,birds, mammals, abundance and distribution, live coral, sea grass cover, microbe biomass and diversity(emerging) and invertebrate abundance and distribution (emerging), 2 DOOS specific EOVs (litter including microplastics, seafloor sponge habitat cover) and 9 MSFD Descriptors (D1, D3, D4, D5, D6, D7, D9, D10, D11) and the following table presents the mapping of a such multitude of capabilities.

Table 1. NAUTILOS Technologies and reference networks

	NAUTILOS Marine Technologies	Task	Variables targeted	Target disciplinary groups	MSFD Descriptor	EMODnet theme	Copernicus Marine Service	EuroGOOS TT	GOOS - International Network
1	Dissolved Oxygen Sensors	(ref. Sub-Task 3.1.1 & Sub-Task 3.1.2)	Dissolved oxygen	Marine biogeochemistry Regulatory environmental monitoring	D3 - Population of commercial fish/shell D4 - Elements of Marine food webs D5 - Eutrophication	Chemistry Physics	yes	Argo Fixed Platforms Gliders FerryBox	ARGO OceanSites OceanGliders Go-SHIP
2	Fluorescence Sensor	(ref. Sub-Task 3.1.2)	Chlorophyll-a fluorescence	Marine biology/ecology Regulatory environmental monitoring	D5 - Eutrophication	Chemistry Physics	yes	Argo Fixed Platforms Gliders FerryBox	ARGO OceanSites OceanGliders Go-SHIP
3	Ocean surface multi/hyperspectral and laser induced chlorophyll-a fluorescence sensors and cameras	(ref. Task 3.2)	Sea surface temperature, Laser induced chlorophyll-a fluorescence, Ocean color	Marine biology/ecology Regulatory environmental monitoring Ocean colour community	D4 - Elements of Marine food webs D5 - Eutrophication	Physics Chemistry	yes	FerryBox	Go-SHIP
4	Passive broadband acoustic recording sensor	(ref. Sub-Task 3.3.1)	Marine noise (anthropogenic and natural sources, including marine mammals, sea ice cracking, seismic activity, meteorological sources)	Marine biology/ecology Regulatory environmental monitoring	D11 - Energy and Water noise D1 - Biological diversity	Physics Biology		Fixed Platforms	OceanSites OBIS
5	Passive acoustic event recorder	(ref. Sub-Task 3.3.2)	Marine mammal sound detection (porpoise & dolphin clicks for abundance estimation)	Marine biology/ecology Regulatory environmental monitoring	D1 - Biological diversity	Biology			OBIS
6	Active Acoustic Profiling Sensor	(ref. Task 3.4)	Suspended particle concentration / distribution (zooplankton, microplastics, organic and inorganic sediment)	Marine biology/ecology Marine pollution	D10 - Marine litter D1 - Biological diversity	Chemistry Physics Biology			OceanSites OBIS
7	Sampler for phytoplankton and other suspended matter	(ref. Task 3.5)	Concentrated suspended matter samplers for analyses of phyto-pigments, particulate organic matter, microbe biomass and diversity	Marine biology Marine ecological monitoring Climate research Marine pollution	D1 - Biological diversity D5 - Eutrophication	Biology Chemistry Physics		Gliders FerryBox	OBIS GO-SHIP
8	Carbonate system/ocean acidification sensors	(ref. Task 4.1)	pH, pCO2,Total Alkalinity	Marine biogeochemistry Climate research	D1 - Biological diversity	Chemistry	yes		GOA-ON/SOCAT

9	Silicate Electrochemical Sensor	(ref. Task 4.2)	Silicate concentration (Si)	Marine biogeochemistry	D5 - Eutrophication	Chemistry		FerryBox	
10	Submersible Nano- and Microplastics Sampler	(ref. Task 4.3)	Concentrated suspended matter samples	Marine ecology Marine pollution	D10 - Marine litter	Physics		FerryBox	
11	Low-cost Microplastic sensors	(ref. Task 4.4)	Concentration and characterisation of microplastics	Marine ecology Marine pollution	D10 - Marine litter	Chemistry		FerryBox	
12	Deep Ocean CTD	(ref. Task 4.5)	Conductivity, Temperature, Pressure (Salinity and Density derived)	Physical oceanography	D7 - Alteration of hydrographical	Physics	yes	Fixed Platforms	DOOS Go-SHIP
13	Deep ocean low-level radioactivity sensor	(ref. Task 4.6)	Radon gas, potassium 40K, radium 226Ra and 228Ra, and other natural isotopes	Environmental monitoring	D9 - Contaminants in fish and seafood	-		Fixed Platforms	DOOS
14	Integration of existing technologies in animal tagging systems	(ref. Task 5.5)	Temperature, Salinity, Chlorophyll-a fluorescence, Dissolved oxygen	Physical oceanography Marine biology/ecology	D7 - Alteration of hydrographical	Physics	yes	AnimalBourne Instrument	AniBOS
15	Demonstration of novel equipment for key seabed habitat mapping	(ref. Task 6.2)	Live corals, hard corals, seafloor sponges	Marine biology	D6 - Seafloor integrity	Seabed habitats			
16	Smartphone NIR Scanner	(citizen science application, ref. Task 8.4)	Plastics	Marine pollution	D10 - Marine litter	Chemistry		FerryBox	
17	Visual marine image annotation	(citizen science application, ref. Task 10.4)	Macroplastics, Sponge and cold-water coral cover, major seafloor organism types.	Marine pollution Marine biology/ecology	D10 - Marine litter	Chemistry		FerryBox	

3. NAUTILOS OBPS REFERENCES

As anticipated the goal of OBPS is to implement procedures for NAUTILOS EOVs, including in information about the sensors and procedures used to measure the variables. NAUTILOS has to refer to common standards and procedures, anyhow for the new sensors, if needed, NAUTILOS has to include the publication of written manuals for the QA/QC procedures. As identified in Task 8.1., these are documents that are part of the outputs of the project to be captured in an open access format and made immediately available to all project partners, also considering both the management procedures (D1.1) and the quality management (D1.4) flows. See Annex 2 for details.

This section provides the reader a quick overview (links) to available documentations and it takes advantages from some closed or running projects that are also addressing the same issues.

AtlantOS[7], which goal was to improve ocean observation capabilities and data sharing, and in particular the WP7 (Data Management), which was involving and networking many ocean-observing platform network, identified a core of seven (common across the networks) EOVs for implementation of common QC procedures (to be up taken by EuroGOOS DATAMEQ).

More specifically the following EOVs

- Physics: temperature (T), Salinity (T), Current for surface and subsurface and Sea level
- Biogeochemistry: Oxygen (O2), Chlorophyll-A, Nitrate (NO3) and Carbon (pCO2) for surface and subsurface

The recommendations[8] have been compiled by experts on those EOVs and validated by the Networks acquiring those EOVs and performing NRT QC. These include both Quality flags that can be mapped to the SeaDataNet flag scale (Figure 2) and when known processing level information ("qualified in NRT using automated procedures" or "processed in DM by Scientist"). These recommendations apply to:

- Argo, Gliders and OceanSITES for T&S, Current, Oxygen(O2), Chla, NO3
- Drifters for T&S, Current
- Ferrybox, FOS, Fixed Moorings, coastal profilers for T&S, Current, Chla
- VOS/SOOP, GO-SHIP for T&S, Chla, Carbon (pCO2), NO3

The document is available at the following link https://atlantos-h2020.eu/download/7.2-QC-Report.pdf

[7] AtlantOS project (https://atlantos-h2020.eu/) goal was to improve and innovate Atlantic observing by using the Framework of Ocean Observing to obtain an international, more sustainable, more efficient, more integrated, and fit-for-purpose system,

[8] https://atlantos-h2020.eu/download/7.2-QC-Report.pdf

Key	Term	Term abbr	Term def
0	no quality control	none	No quality control procedures have been applied to the data value. This is the initial status for all data values entering the working archive.
1	good value	good	Good quality data value that is not part of any identified malfunction and has been verified as consistent with real phenomena during the quality control process.
2	probably good value	probably good	Data value that is probably consistent with real phenomena but this is unconfirmed or data value forming part of a malfunction that is considered too small to affect the overall quality of the data object of which it is a part.
3	probably bad value	probably bad	Data value recognised as unusual during quality control that forms part of a feature that is probably inconsistent with real phenomena.
4	bad value	bad	An obviously erroneous data value.
5	changed value	changed	Data value adjusted during quality control. Best practice strongly recommends that the value before the change be preserved in the data or its accompanying metadata.
6	value below detection	BD	The level of the measured phenomenon was too small to be quantified by the technique employed to measure it. The accompanying value is the detection limit for the technique or zero if that value is unknown.
7	value in excess	excess	The level of the measured phenomenon was too large to be quantified by the technique employed to measure it. The accompanying value is the measurement limit for the technique.
8	interpolated value	interpolated	This value has been derived by interpolation from other values in the data object.
9	missing value	missing	The data value is missing. Any accompanying value will be a magic number representing absent data.
A	value phenomenon uncertain	ID_uncertain	There is uncertainty in the description of the measured phenomenon associated with the value such as chemical species or biological entity.

Figure 2. SDN Quality Flags and definitions

Table 2. Best Practices procedures for NAUTILOS tools

Document Name	Link	Applies to:						
		FOS	FB	Gliders	AUVs	ABI	ARGO	MO and FP
AtlantOS - Handbook for Data Management	https://archimer.ifremer.fr/doc/00370/48139/48242.pdf	yes	yes	yes	Yes	Yes	yes	Yes
ARGO User's Manual	https://archimer.ifremer.fr/doc/00187/29825/86414.pdf	-	-	-	-	-	yes	-
OceanSITES User's Manual	http://www.oceansites.org/docs/oceansites_user_manual_ver1_1.pdf	-	-	-	-	-	-	yes
EuroGOOS reccomendations of In SITU data NRT QC	https://repository.oceanbestpractices.org/handle/11329/656	-	yes	yes	Yes	-	yes	yes
SeaDataCloud - Manaual for Flow Cytometry Data ingestion, validation and long term storage	https://www.seadatanet.org/content/download/3721/file/SDC_WP9_D9.13_FlowCytometryDataManagement.pdf	yes	Yes	-	-	-	-	-
OceanGlider Data Format	https://github.com/OceanGlidersCommunity/OG-format-user-manual/blob/main/OG_Format.adoc	-	-	yes	Yes	-	-	-

OceanSITES Data Format	http://www.oceansites.org/docs/oceansites_data_format_reference_manual.pdf	-	-	-	-	-	-	Yes
Ferrybox Data Format	https://www.ferrybox.com/imperia/md/content/ferryboxusergroup/ferrybox_d-3-3-b__data_management_guidelines__r_2-0.pdf	-	Yes	-	-	-	-	-
MEOP Data Format	https://www.meop.net/database/format/	-	-	-	-	yes	-	-
Marine Litter User	https://doi.org/10.13120/21addf37-7e82-4a55-b040-3d3d87115ac0	-	-	-	-	-	-	-
Guidelines and formats for gathering and management of micro-litter data sets on a European scale	https://doi.org/10.6092/d3e239ec-f790-4ee4-9bb4-c32ef39b426d	-	-	-	-	-	-	-
Marine Litter Technical Recommendations for the Implementation of MSFD Requirements	https://mcc.jrc.ec.europa.eu/documents/201702071118.pdf	Yes	Yes	-	-	-	-	-
Cruise Summary Report	https://www.seadatanet.org/content/download/7287/file/sdn_csr_backoffice_usermanual_V2.0.pdf	Yes	Yes	-	-	-	-	-

Animal Borne Instrument and Animal Telemetry	https://www.frontiersin.org/articles/10.3389/fmars.2019.00326/full	-	-	-	-	yes	-	-
Glider Community github	https://github.com/OceanGliders Community	-	-	yes	yes	-	-	-

4. PROMOTING THE USE OF NAUTILOS DATA

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.

While the data flow for classical ocean physical parameters such as temperature and salinity, as described in the previous section, is well-defined, 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.

One example to present the general case is “Marine litter”.

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.

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.

To describe the litter, EMODnet Chemistry has developed a document about “Guidelines and formats for gathering and management of micro-litter data sets on a European scale”. Eventually, whatever is the methodology to collect the information, to be immediately

consumable by EMODnet Chemistry litter registry, the information has to be described in a standard metadata format, i.e. Common Data Index (CDI)[9]

5. NAUTILOS REFERENCE VOCABULARIES

As described in the DMP, NAUTILOS is adopting the following vocabularies.

Table 3. Metadata Vocabularies

Metadata field	**Vocabulary exists**	**Link to vocabulary**	**Vocabulary governance**
Platform type	Yes	http://vocab.nerc.ac.uk/collection/L06/current/	BODC
Platform type bigram	Yes	CMEMS INSTAC	EuroGOOS DATAMEQ
sensor_model	Yes	http://vocab.nerc.ac.uk/collection/L22/current/	BODC -NVS
ICES_code	Yes	https://ocean.ices.dk/codes/ShipCodes.aspx	ICES
contributors_role			NAUTILOS
naming_authority	Yes	https://edmo.seadatanet.org/	SeaDataNet
Institution	Yes	https://edmo.seadatanet.org/	SeaDataNet
qc_method	*	doi	
data_mode	Yes	RT/DM/REP	EuroGOOS DATAMEQ
Phase	No		NAUTILOS
variable names	Yes	http://vocab.nerc.ac.uk/collection/P02/current/ http://vocab.nerc.ac.uk/collection/P01/current/ http://vocab.nerc.ac.uk/collection/P07/current/	BODC - NVS
Time	yes	ISO8601	ISO
Datum	Yes	WGS84	ISO
Taxon	Yes	LSID	WoRMS
Country	yes	ISO3166	ISO
Licence	Yes	https://creativecommons.org/	CC
INSPIRE	Yes	ISO 19115	ISO/INSPIRE

*) RTQC are usually described in scientific papers, here the recommendation is to include the “doi” of such paper-methodology.

[9] https://www.emodnet-chemistry.eu/repository/Proposal-EMODnet-TG-ML-Micro-Litter-Data-Gathering-03062020.pdf

6. IMPROVING GLOBAL ATTRIBUTES

As described in the previous sections, according the parameter and the recording platform, NAUTILOS can refer to and adapt already widely adopted standards and best practices. The same analysis highlighted that a key element in the ocean best practice is the proper compilation of metadata.
Usually, this information is reported in the global attributes, but it may not be valid for all the parameters in the data model. Hence here NAUTILOS propose to move this information at the level of the parameter where it will be possible to refer to different QC/QF for different parameters.

The following table shows an example of a subset of the global attributes for a CTD file and in yellow the reference to the QC/QF methodology

Row Type	Variable Name	Attribute Name	Data Type	Value
attribute	NC_GLOBAL	_NCProperties	String	version=1|netcdflibversion=4.4.1|hdf5libversion=1.8.18
attribute	NC_GLOBAL	area	String	Arctic Ocean
attribute	NC_GLOBAL	cdm_data_type	String	Profile
attribute	NC_GLOBAL	cdm_profile_variables	String	PLATFORMCODE,time,latitude,longitude
attribute	NC_GLOBAL	citation	String	These data were collected and made freely available by the Copernicus project and the programs that contribute to it
attribute	NC_GLOBAL	contact	String	cmems-service at imr.no
attribute	NC_GLOBAL	Conventions	String	CF-1.6 OceanSITES-Manual-1.2 Copernicus-InSituTAC-SRD-1.4 Copernicus-InSituTAC-ParametersList-3.1.0, COARDS, ACDD-1.3
attribute	NC_GLOBAL	creator_email	String	katrin.schroeder at ismar.cnr.it
attribute	NC_GLOBAL	creator_name	String	Katrin Schroeder
attribute	NC_GLOBAL	creator_type	String	person
attribute	NC_GLOBAL	creator_url	String	http://www.oceansites.org
attribute	NC_GLOBAL	data_assembly_center	String	IMR
attribute	NC_GLOBAL	data_mode	String	R
attribute	NC_GLOBAL	data_type	String	OceanSITES vertical profile
attribute	NC_GLOBAL	date_update	String	2021-10-21T11:42:16Z
attribute	NC_GLOBAL	distribution_statement	String	These data follow Copernicus standards; they are public and free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data. User must contact PI prior to any commercial use of data.
attribute	NC_GLOBAL	Easternmost_Easting	double	179.9978
attribute	NC_GLOBAL	featureType	String	Profile
attribute	NC_GLOBAL	format_version	String	1.4
attribute	NC_GLOBAL	geospatial_lat_max	double	89.9909
attribute	NC_GLOBAL	geospatial_lat_min	double	-78.35756
attribute	NC_GLOBAL	geospatial_lat_units	String	degrees_north
attribute	NC_GLOBAL	geospatial_lon_max	double	179.9978
attribute	NC_GLOBAL	geospatial_lon_min	double	-179.9871

attribute	NC_GLOBAL	geospatial_lon_units	String	degrees_east
attribute	NC_GLOBAL	geospatial_vertical_max	double	5395.0
attribute	NC_GLOBAL	geospatial_vertical_min	double	0.1
attribute	NC_GLOBAL	geospatial_vertical_positive	String	down
attribute	NC_GLOBAL	geospatial_vertical_units	String	m
attribute	NC_GLOBAL	id	String	AR_PR_CT_ITP-103_201906
attribute	NC_GLOBAL	infoUrl	String	http://www.oceansites.org
attribute	NC_GLOBAL	institution	String	institution name
attribute	NC_GLOBAL	institution_country	String	United States
attribute	NC_GLOBAL	institution_edmo_code	String	3844
attribute	NC_GLOBAL	institution_references	String	https://www.whoi.edu/itp
attribute	NC_GLOBAL	keywords	String	cnr, cnr-iamc, consiglio, costiero, council, data, DATA_MODE, delayed, delle, density, direction, earth, Earth Science > Oceans > Ocean Pressure > Water Pressure, Earth Science > Oceans > Ocean Temperature > Potential Temperature, Earth Science > Oceans > Salinity/Density > Salinity, flag, iamc, institution, istituto, italy, l'ambiente, latitude, longitude, marino, med, mode, name, national, nazionale, near, nrt, observations, ocean, oceans, per, potential, POTENTIAL_TEMP, POTENTIAL_TEMP_QC, practical, PRES, PRES_QC, pressure, profiles, PSAL, PSAL_QC, quality, real, research, ricerche, salinity, science, sea, sea_water_potential_temperature, sea_water_practical_salinity, sea_water_pressure, seawater, situ, temperature, time, TIME_QC, title, water
attribute	NC_GLOBAL	keywords_vocabulary	String	GCMD Science Keywords
attribute	NC_GLOBAL	last_date_observation	String	2019-06-30T21:02:00Z
attribute	NC_GLOBAL	last_latitude_observation	String	77.04270
attribute	NC_GLOBAL	last_longitude_observation	String	-129.78680
attribute	NC_GLOBAL	license	String	These data follow Copernicus standards; they are public and free of charge. User assumes all risk for use of data. User must display citation in any publication or product using data. User must contact PI prior to any commercial use of data.
attribute	NC_GLOBAL	naming_authority	String	Copernicus Marine In Situ
attribute	NC_GLOBAL	netcdf_version	String	netCDF-4 classic model
attribute	NC_GLOBAL	Northernmost_Northing	double	89.9909
attribute	NC_GLOBAL	pi_name	String	John M. Toole
attribute	NC_GLOBAL	platform_code	String	ITP-103
attribute	NC_GLOBAL	platform_name	String	ITP 103
attribute	NC_GLOBAL	qc_manual	String	Recommendations for in-situ data Near Real Time Quality Control https://doi.org/10.13155/36230
attribute	NC_GLOBAL	quality_control_indicator	int	1
attribute	NC_GLOBAL	quality_index	String	A
attribute	NC_GLOBAL	references	String	http://marine.copernicus.eu http://www.marineinsitu.eu
attribute	NC_GLOBAL	source	String	Ice-tethered subsurface profiling float
attribute	NC_GLOBAL	source_platform_category_code	String	4A
attribute	NC_GLOBAL	sourceUrl	String	(local files)

attribute	NC_GLOBAL	Southernmost_Northing	double	-78.35756
attribute	NC_GLOBAL	standard_name_vocabulary	String	CF Standard Name Table v55
attribute	NC_GLOBAL	subsetVariables	String	PLATFORMCODE
attribute	NC_GLOBAL	summary	String	CTD in situ Observations
attribute	NC_GLOBAL	time_coverage_end	String	2022-02-22T01:59:33Z
attribute	NC_GLOBAL	time_coverage_start	String	2014-01-01T00:00:01Z
attribute	NC_GLOBAL	title	String	CTD in situ Observations
attribute	NC_GLOBAL	update_interval	String	void
attribute	NC_GLOBAL	Westernmost_Easting	double	-179.9871
variable	PLATFORMCODE		String	

And an example for a parameter:

variable	TEMP		float	
attribute	TEMP	_ChunkSizes	int	570, 377
attribute	TEMP	_FillValue	float	NaN
attribute	TEMP	actual_range	float	-46.879, 815.861
attribute	TEMP	ancillary_variables	String	TEMP_QC TEMP_DM
attribute	TEMP	colorBarMaximum	double	32.0
attribute	TEMP	colorBarMinimum	double	0.0
attribute	TEMP	data_mode	String	R
attribute	TEMP	long_name	String	Sea temperature
attribute	TEMP	standard_name	String	sea_water_temperature
attribute	TEMP	units	String	degree_C
attribute	TEMP	sdn_parameter_urn	String	SDN:P01::TEMPST01
attribute	TEMP	valid_max	float	40.0
attribute	TEMP	valid_min	float	-2.5
variable	TEMP_QC		byte	
attribute	TEMP_QC	_ChunkSizes	int	570, 377
attribute	TEMP_QC	_FillValue	byte	127
attribute	TEMP_QC	actual_range	byte	0, 9
attribute	TEMP_QC	colorBarMaximum	double	10.0
attribute	TEMP_QC	colorBarMinimum	double	0.0
attribute	TEMP_QC	conventions	String	Copernicus Marine In Situ reference table 2
attribute	TEMP_QC	flag_meanings	String	no_qc_performed good_data probably_good_data bad_data_that_are_potentially_correctable bad_data value_changed value_below_detection nominal_value interpolated_value missing_value
attribute	TEMP_QC	flag_values	byte	0, 1, 2, 3, 4, 5, 6, 7, 8, 9
attribute	TEMP_QC	long_name	String	Sea temperature quality flag
attribute	TEMP_QC	valid_max	byte	9
attribute	TEMP_QC	valid_min	byte	0

The recommendation is then to move the reference of the applied QC/QF best practice at the level of the parameter attributes e.g.:

attribute	TEMP_QC	QC_method	String	https://repository.oceanbestpractices.org/handle/11329/656

7. OPEN ACCESS DOCUMENTATION WORKFLOW

To define the NAUTILOS document management workflow, different tools and methodologies were considered (Appendix 2) and the identified workflow is based on Google Team Drive and OwnCloud to be used as follow:

- Each WP representative have access to specific directories on the ownCloud platform where templates are uploaded and stored when finished
- Documents templates are then accessible for being downloaded and subsequentially uploaded on the collaborative elaboration platform (Google Team Drive) were all the involved WP members can work simultaneously
- Once a document is ready and validated is then uploaded on the ownCloud delivery directory in pdf

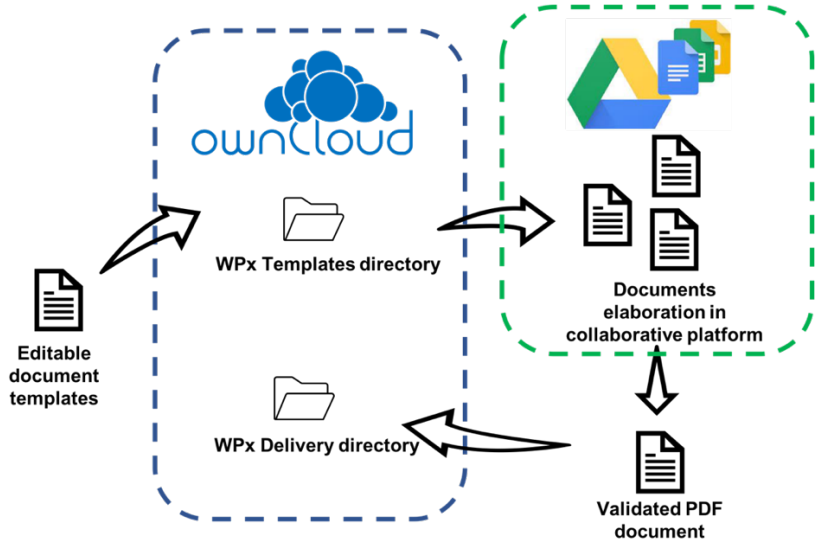

Figure 3: Schema of the Document flow

Once the document is validated, in order to have a wider distribution, access and being able to track the impact of the NAUTILOS outcome the document has to be published with a digital object identifier doi. NAUTILOS recommends to use Zenodo as a final repository for NAUTILOS documentation. Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN. It allows to deposit research papers, data sets, research software, reports, and any other research related digital artefacts. Interestingly Zenodo offers the possibility of reserve the DOI, in this way the DOI is not formally generated and data are not published yet allowing modifications and updating data. Once the record has been published, files can no longer be modified but the system supports file versioning hence there is the possibility to updated a document by keeping the same doi root.

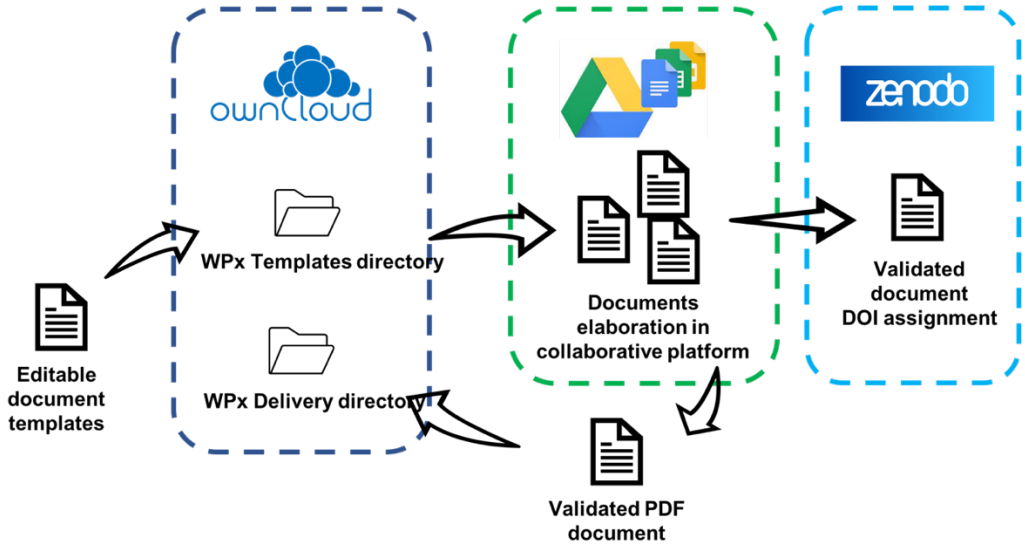

Figure 4: Schema of the Document flow

8. OTHER USEFUL OBPS REFERENCES

FerryBox Whitebook	https://repository.oceanbestpractices.org/handle/11329/1502
Climate and Forecast Metadata Convention CF-1.8	http://cfconventions.org/Data/cf-conventions/cf-conventions-1.8/cf-conventions.html#trajectory-data
Unidata NetCDF Attribute Convention for Data Discovery (ACDD)	https://www.unidata.ucar.edu/software/netcdf-java/v4.6/metadata/DataDiscoveryAttConvention.html
Maritime Sensor Technologies for the European Market: Research, Development and Implementation. Good practice guide.	https://repository.oceanbestpractices.org/handle/11329/431
QARTOD	https://ioos.noaa.gov/project/qartod/

9. Appendix 1: References and Related Documents

ID	Reference or Related Document	Source or Link/Location
1	NAUTILOS DMP	10.5281/zenodo.7163625
2	NAUTILOS DATA QUALITY DOCUMENT	10.5281/zenodo.7163673

10. APPENDIX 2

To identify the workflow and the deployable document management system for the NAUTILOS project the following tools were evaluated:

- AlFresco is a collection of products for information management developed for Windows and Unix operative systems, it is based on Java technology. The primary software, Digital Business Platform, is a licensed product that support open-source platform. A free Community Relies is also provided with functional limitations. It is made for modular and scalable performances, it could be deployed on server or in cloud using AWS
- ownCloud is a client-server software, his functionalities allow to store files in a traditional folder structure. It is devolved for Windows, OS and Linux operative systems. At management level this system allows to control users or groups permissions (read/write) to high granular level. It also provides business solutions allowing additional functionalities and support services
- SharePoint is a web-based collaborative platform integrated with Microsoft Office products. Its functionalities allow a high personalization providing a wide range of possible usages. This software allows real-time documents sharing and editing with the addition of integrated features for graphic data visualization. It can be purchased separately or integrated in Office 365
- Google Tema Drive is part of the web-based platform Google Workspace and allow files share and synchronisation. It is accessible on online platform and offline application for Windows and OS operative systems. It is integrated with Google Sheets, Docs and Slide that are part of the Google Office Suite and allow real-time sharing and editing. The product offer both open and licensed solutions with differences in storage capacity and advanced functionalities

The assessment considered the following performance indicators:

- Usability, evaluate the easiness of deployment, user interface and hardware requirements.
- Portability, evaluate the integration of the product on diverse operative systems taking into account both online and offline possibility.
- Costs, evaluate the cost of license and additional features needed for the system deployments.
- Storage capacity, evaluate the maximum storage capacity.

The following table presents the results of the analysis.

Table 4: Shared working space providers evaluation

	Usability					Portability			Cost		Storage Capacity
	Deployment on server	Programming needed	User Friendly	Additional software needed	Collaborative working	Work on Unix-based and windows	Online platform	Offline documents update	Open Sucre	Licenses cost	Storage capacity
AlFresco	X	X	X			X					
ownCloud			X	X		X		X	X		
SharePoint	X	X		X	X		X	X			
Google Team Drive			X		X	X	X	X	X		

The analysis highlighted Google Team Drive to be the most suitable system for documents collaborative work and share. It combines a real-time collaborative documents compilation environment (Google Office Suite) and an easy sharing files system, with a high portability on any operative systems as well as on mobile systems. Google Team Drive also supports Microsoft Office and Open Office files and formats.