



Report and release of IT-IOOS metadata catalogue of data and facilities



Deliverable number:	D5.11
Work package:	WP5 – Marine Domain
Intermediate Objective:	IO5.5
Deliverable type:	<input checked="" type="checkbox"/> Document, report
	<input type="checkbox"/> Websites, patent filings, videos, etc.
	<input type="checkbox"/> Other: please specify
Dissemination level:	<input checked="" type="checkbox"/> Public
	<input type="checkbox"/> Restricted
Estimated delivery (bimester):	B12
Actual delivery date:	30/09/2024
Author(s) (Partner-OU):	Alessandra Giorgetti (OGS), A. Cociancich (OGS), E. Partescano (OGS)
Reviewed by:	M. Caccavale (CNR-ISMAR), A. Fadini (CNR-ISMAR), V. Langella (CNR-ISMAR), ITINERIS Executive Board
Note:	Update after the meeting with WP2 to align IT-IOOS catalogue of the marine domain with the ITINERIS catalogue, July 2024

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006 (D.D. n. 130/2022)
 Funded by EU - Next Generation EU
 Mission 4 “Education and Research” - Component 2: “From research to business” -
 Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures”

Table of contents

<i>1. INTRODUCTION</i>	<i>4</i>
1.1. Purpose of the document	4
1.2. IT-IOOS data and facilities	4
1.3. The facilities	5
1.4. What are metadata and why are they so important?	5
1.5. Metadata and FAIRness in the International framework	6
1.6. Metadata and FAIRness in the National framework	9
1.7. Applicable and reference documents	10
1.8. Glossary, Acronyms and Abbreviations	11
<i>2. KEY FEATURES AND COMPONENTS</i>	<i>13</i>
2.1. The definition of CKAN dataset	13
2.2. IOOS metadata catalogue	13
2.2.3. Metadata of the Sub-System	14
2.2.4. Metadata of the Facility or Observation Platform	14
2.2.5. Metadata of the Dataset	15
<i>3. INITIAL IMPLEMENTATION AND MAINTENANCE STRATEGY</i>	<i>16</i>
3.1. Short-term Strategy	16
3.1.3. Initial Implementation.....	16
3.1.4. Operational Support	17
3.2. Long-term Strategy	17
3.3. Alignment with the ITINERIS Catalogue of data and services	17
3.4. Alignment with SeaDataNet Catalogues	18
<i>4. ANNEX - METADATA CATALOGUE FORM</i>	<i>19</i>

1. INTRODUCTION

1.1. Purpose of the document

The aim of this report is to document the release of the metadata catalogue for data and facilities related to the Italian Integrated Ocean Observing System (IT-IOOS). It represents a critical component of the IT-IOOS architecture and one of the first elements of the implementation plan, as outlined in Deliverable 5.6: Report on the Design of the Italian Integrated Ocean Observing System (IT-IOOS). In particular, the IT-IOOS metadata catalogue provides a comprehensive description of all components of the sub-systems, their data and products. It serves as an important link to federate all data and data products provided by the marine research infrastructures (RIs) of ITINERIS, the Marine Hub (comprehensive of the Data Portal and the Marine Data Store). The short and long-term strategy for the governance and maintenance of the IT-IOOS metadata catalogue is also proposed.

The release of the IT-IOOS metadata catalogue represents an important milestone in the development of the Italian Integrated Ocean Observing System. The provision of a centralized and standardized metadata repository will enable better integration and use of marine data and products across different research infrastructures. The proposed governance and maintenance strategies will ensure the continued relevance and effectiveness of the metadata catalogue in supporting marine research and observation both within the ITINERIS project period and in the coming decade.

This report is to be regarded as a living document that will be updated during the process of implementing and compiling the IT-IOOS catalogue.

1.2. IT-IOOS data and facilities

The Italian Integrated Ocean Observing System (IT-IOOS) builds on several national research infrastructure nodes identified in the National Plan for Research Infrastructures 2021-2027. Some of them are also supported by the Research Infrastructures Programme and the European Strategy Forum on Research Infrastructures (ESFRI) established by the European Commission in collaboration with the Member States and the scientific community. The ITINERIS project has provided the opportunity to harmonize the Italian marine research infrastructures, including their data, products, facilities, and services.

In particular, IT-IOOS consists of the following the Sub-Systems (SS):

1. DANUBIUS
2. eLTER
3. EMSO
4. EuroARGO
5. EuroFleet
6. GEOSCIENCE-MER
7. ICOS
8. JERICO
9. Laura Bassi
10. OCEAN SOUND
11. SIOS
12. Marine Data Store
13. Marine Data Portal

The architecture, the main components and the modules of the various SS within IT-IOOS are defined in section 6 of Deliverable 5.6, together with the functions of the Production Units (PU), which have the task of generating and making data and products accessible.

1.3. The facilities

The facilities of the IT-IOOS sub-systems include various types of in situ observation platforms such as buoys, research vessels, and remote sensing instruments. These can collect physical, chemical, and biological data on the marine environment. Besides, reanalysis and model data can be also generated on demand by a few sub-systems. Each facility is fully described in the metadata catalogue, with all its instruments, sensors, and all related data and data products.

1.4. What are metadata and why are they so important?

Metadata is essentially "data about data". It provides crucial information to describe the data, its consistency, its origin. Metadata replies to the questions of **what** data we are dealing with (data content and quality), **who** originated the data, **how** the data was generated, **where** and **when** it was generated. Metadata represent the indispensable part of information that characterizes the data. Structured and standardized metadata are fundamental in the description and management of the data. Metadata helps to identify and discovery the data.

Indeed, structured and standardized metadata are crucial for achieving the principles of FAIRness (Findable, Accessible, Interoperable, and Reusable) and Openness, particularly in the context of marine data, where datasets can be vast and diverse (ranging from satellite observations to biological and chemical measurements). Structured metadata is especially important for promoting global collaboration, advancing research and policymakers in ocean science and environmental management.

Metadata also describes how the data is organised and structured, including formats, file types and relationships between different data elements, and contains information needed to manage the data, such as creation and modification dates, access rights and ownership.

Metadata is important for several reasons:

1. For data discovery and access: metadata makes it easier for users to find and access data. Descriptive metadata allows researchers to search for datasets based on keywords, author names or other relevant information. By providing context and summaries, metadata helps users quickly determine the relevance of a dataset to their needs.
2. For data organisation and management: describing the structure and relationships of datasets is essential for efficient storage, retrieval and integration of data.
3. For data governance: by specifying access rights, usage restrictions and the origin of the data.
4. Data quality: metadata contains information about the quality and source of the data and helps users to assess its reliability and suitability for their purposes.
5. Data interoperability: Standardised metadata formats enable interoperability between different systems and datasets and facilitate the sharing and integration of data across different platforms and institutions.

For data preservation and sustainability: metadata ensures long-term preservation by documenting the essential characteristics and context of a dataset. This is crucial for future access and use, even if the original context or technology changes.

6. To support sustainability by providing information needed to maintain, update and migrate datasets over time.
7. For improving research and collaboration: by providing comprehensive descriptions and contextual information, metadata enables better understanding and use of data, fostering collaboration

between researchers. They support the reproducibility of research by documenting in detail how the data was collected, processed and analysed.

Metadata is a fundamental element of data management and utilisation. It improves the findability, accessibility, quality and sustainability of data, making it an indispensable tool for researchers, data managers and institutions. By using metadata effectively, organisations can ensure that their data assets are well organised, easily accessible and valuable to a wide range of applications and users.

1.5. Metadata and FAIRness in the International framework

Structured and standardized metadata, on the data itself and on the semantic content, are essential for achieving FAIRness and Openness of (marine) data.

In particular, the EU project ENVRI-FAIR (ENVironmental Research Infrastructures building Fair services Accessible for society, innovation and research) aimed to advance the findability, accessibility, interoperability, and reusability (FAIRness) of the data and services offered by the ENVRI Cluster of research infrastructures and to connect them to the emerging European Open Science Cloud. Environmental data and services produced by the different research infrastructures were harmonized with a holistic approach and made easy to use for scientists from any field of environmental research. Five infrastructures contributed to the Marine domain, namely EuroARGO, EMSO, ICOS-Ocean, Life-Watch and SeaDataNet, who participated with the aim of contributing their expertise and potential solutions, as well as seeking closer co-operation and networking with the marine RIs.

SeaDataNet is the pan-European Infrastructure for Ocean & Marine data management and was established in early 2000 by a funded European project to provide a framework for data sharing and co-operation. It provides a single website access to multidisciplinary ocean and marine data as well as data products and metadata services. SeaDataNet involves more than 110 data centres and it makes available more than 2 million datasets acquired from research cruises, marine infrastructures, and other observational activities in European and global waters. Over the years, SeaDataNet has defined de-facto standards for data, metadata and vocabularies largely used these days, collaborating with European and international experts within the framework of IOC-IODE, ICES, adapting ISO and OGC standards, and achieving INSPIRE compliance for some metadata services directory. SeaDataNet serves the ocean and marine community as well as government agencies, industry, researchers and the general public, allowing it to preserve ocean data over a long period, visualise and analyse it. SeaDataNet provides access to extremely heterogeneous marine and ocean data in an effective and easy way.

The SeaDataNet metadata services give overviews of marine organisations in Europe and their engagement in marine research projects, managing large datasets, and data acquisition by research vessels and monitoring programmes for the European seas and global oceans.

In SeaDataNet the metadata are organised over five catalogues, listed in the following:

- CSR = Cruise Summary Reports
- EDMED = European Directory of Marine Environmental Data
- EDMERP = European Directory of Marine Environmental Research Projects
- EDIOS = European Directory of Initial Ocean-observing Systems
- EDMO = European Directory of Marine Organisations

These metadata catalogues are harmonised in the use of syntax, semantics and tools. The maintenance of these directories - key service for unified access to data sets - is coordinated at a national level by the National Oceanographic Data Centres / SeaDataNet Marine Data Centres, that collect references from institutes in their country and enter these national entries into the pan-European directories.

All SeaDataNet metadata services make use of XML formats, based upon the ISO 19115 content model, exchange schema's, based upon the ISO 19115 DTD, and Controlled Vocabularies, which underpins the relations and semantic coherence between the different metadata services. However, for the CSR

and CDI services an upgrade to ISO 19139 format has been applied, especially for making these INSPIRE compliant and for making the exchange between the national providers and the central portal operators more efficient by automatic harvesting. SeaDataNet makes available also an important tool, MIKADO software, that can be used to enter and maintain local directories as sets of XML files.

In addition, the CDI = Common Data Index system connects the SeaDataNet portal and the databases with the datasets of the SeaDataNet distributed National Oceanographic Data Centres / SeaDataNet Marine Data Centres. The maintenance of the CDI system is undertaken by the distributed data centres that operate local and central components of the CDI system. The CDI service delivers datasets to users in standard SeaDataNet Data Transport Formats. This implies that the distributed data centres must pre-process data via a conversion routine and arrange that the data files are available in the SeaDataNet formats. Software tools (namely NEMO and OCTOPUS) are developed by SeaDataNet for data preparation.

The SeaDataNet metadata directories (CDI, CSR, EDMED, EDMERP, EDIOS and EDMO) content can also be accessed through various machine-to-machine interfaces (CDI API, SPARQL Endpoint and WMS and WFS services): <https://cdi.seadatanet.org/#m2m-interface>.

Controlled vocabularies have been introduced to standardize how data and information are exposed to the web, as the content has to be machine readable. In the case of SeaDataNet, the vocabularies are audited and published on the NERC Vocabulary Server (NVS 2.0) in the machine-readable, Simple Knowledge Organization System (SKOS) with standardized APIs for querying and delivering terms (REST, SOAP and SPARQL). Many of these vocabularies are also semantically linked to local or external vocabularies, so a user (or machine) can identify similar or related terms.

The IT-IOOS catalogue is designed according to the SeaDataNet standard to enable easy integration into the European network. In particular, the metadata fields of the IT-IOOS catalogue are derived from the EDIOS catalogue and the CDI. In addition, the IT-IOOS catalogue uses controlled vocabularies. These elements facilitate the integration of IT-IOOS information into SeaDataNet.

At the international level, beyond the European context, several integrated ocean observing systems have been developed, characterized by their adherence to FAIR principles and supported by comprehensive metadata systems. Examples include the Integrated Ocean Observing System (IOOS) under NOAA (U.S.), the Canadian Integrated Ocean Observing System (CIOOS), the Australian Ocean Data Network (AODN) and the Japan Oceanographic Data Center (JODC).

The Integrated Ocean Observing System (part of NOAA - National Oceanic and Atmospheric Administration) is a national-regional partnership working to provide new tools and forecasts to improve safety, enhance the economy, and protect our environment. Integrated ocean information is available in near real time, as well as retrospectively. Easier and better access to this information is improving the ability to understand and predict coastal events - such as storms, wave heights, and sea level change. Such knowledge is needed for everything from retail to development planning. IOOS collect data from regional partners as: SECOORA, MARACOOS, NERACOOS, GLOS, GCOOS, CARICOOS, SCCOOS, CeNCOOS, NANOOS, PacIOOS, AOOS. All 11 IOOS Regional Associations offer data collected in their region through their website data portals. The data offered at each portal is different and focuses on the regional associations' specific strengths and local partnerships. The quality of data provided by regional partners is guaranteed by certification standard. The IOOS Data Catalog is the master inventory of IOOS DMAC (Data Management and Communications) datasets and data access services. Users of the Catalog can do a plain text search to locate datasets of interest across the regions, and then connect to their associated data access services in their application of choice. The Catalog is powered by metadata and allows users the ability to search for words or phrases within the metadata (eg. salinity observations) to locate relevant datasets. The DMAC provides a guideline to share data through the IOOS data portal.

The Australian Ocean Data Network (AODN) is an interoperable online network of marine and climate data resources. IMOS and the 6 Australian Commonwealth agencies form the core of the AODN. Increasingly, though, universities and State government offices are offering up data resources to the

AODN, and delivery of data to the AODN is being written in to significant research programs e.g. National Environmental Science Program Marine Biodiversity Hub and the Great Australian Bight research program. The marine data collections are made freely available to the public. Data covers a wide range of parameters in different ocean environments collected from ocean-going ships, autonomous vehicles, moorings and other platforms. The scope of observations, geographically spanning ocean to coast, and across disciplines (physical, biogeochemical, biological), provides a challenge to deliver an intuitive easy-to-use robust information infrastructure enabling users to efficiently obtain the data they need. The AODN Portal is a single access point for marine data published by AODN contributors. The Portal incorporates a catalogue of metadata, a search interface driven by facets utilising controlled vocabulary terms, and a map interface that can be used to interact with AODN datasets and offers data download in several formats. All available data collections are listed in the main panel with a map showing an approximate data coverage. All the data can be filtered by Measured parameter, Organisation, Platform, Date, Geographic Boundary and many other parameters based on a controlled vocabulary.

The Canadian Integrated Ocean Observing System (CIOOS) was established in 2019 to address inaccessible ocean data and the lack of a mechanism for coordination and collaboration in Canada's ocean observing community. CIOOS works to integrate fragmented and isolated ocean data by creating an open-access national platform that facilitates data discovery for end-users. CIOOS curates Essential Ocean Variables (EOVs) collected in Canadian waters aligned with the Global Ocean Observing System (GOOS). The CIOOS contribute to global initiatives and are part of Canada's commitment to the United Nations Decade of Ocean Science for Sustainable Development (2021 – 2030). These endeavours strive to make ocean science more accessible, predictable, and engaging for all. CIOOS is a collaboration between institutional, governmental, and non-governmental partners located in the Pacific, the St. Lawrence, and the Atlantic. These Regional Associations (RAs) work closely with local organizations to meet end-user needs and develop place-based solutions for a diverse country.

As the synthetic marine data bank of Japan, the Japan Oceanographic Data Center (JODC) collects and manages the oceanographic data observed by various organizations in Japan, which include the governmental organizations, universities and the other marine research institutes. Then JODC ensures data quality and provides various users with such marine data. JODC has been also acting as National Oceanographic Data Center of International Oceanographic Data and information Exchange (IODE) promoted by UNESCO/IOC since 1965. JODC has archived the oceanographic data and related information obtained by various oceanographic research institutes in and outside Japan and has served the collected data and information for users' activities. J-DOSS is a system to provide users with the data managed by JODC through the Internet. Any user can search the data using many keywords and download them from J-DOSS.

Since the first EU project ODIP (Ocean Data Interoperability Platform) in 2012, several efforts have been made by the SeaDataNet community to enable a more efficient exchange of standards and sharing of information and data across scientific domains and international borders. To this end, all major organisations involved in ocean data management in Europe, the USA, and Australia have joined the ODIP partnership, which is also supported by the IOC/IODE, which participates in the implementation and operation and closely links this activity to its ODSBP project. ODIP provided a platform to assess potential standards and interoperability solutions, harmonising the various regional systems while driving the European contribution to the global system.

Further to the use and importance of standards, standardizing the encoding of metadata and data themselves is crucial if data are to be readily usable by machines or dataset aggregations.

Dedicated tools as the machine-actionable FAIR Implementation Profiles (FIP) can support communities in the decision process leading to optimal FAIR implementations and practices and are available here: <https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>.

F-UJI web service programmatically assesses FAIRness of research data objects at the dataset level based on the FAIR Data Object Assessment Metrics as explained here <https://www.f-uji.net/>.

1.6. Metadata and FAIRness in the National framework

Several environmental RIs, Sub-Systems of IT-IOOS, are funded by the Research Infrastructures Programme and the European Strategy Forum on Research Infrastructures (ESFRI) and have been contributing data and metadata to SeaDataNet for several years.

Through the implementation of ITINERIS and IT-IOOS, Italy will contribute to European and international efforts in the field of ocean observation (European Ocean Observing System, EOOS and Global Ocean Observing System, GOOS) by providing new and qualified data and metadata.

It is therefore crucial to start updating and completing the European metadata directories, maintained by SeaDataNet, with the new ITINERIS efforts. As a first step, all Italian marine research institutes, data holding centres, monitoring agencies, governmental and private organisations involved in ITINERIS in one way or another should be listed in the EDMO. Secondly, all observation platforms (facilities) of the IT-IOOS Sub-Systems should be listed in EDIOS.

As an example, given that all EuroFleets+ cruises are listed in SeaDataNet, having the catalogue enriched with ITINERIS activities, will result in having all metadata easily included in ITINERIS and the two catalogues operationally synchronised.

In the same way, the observing platforms linked to EMSO or JERICO when listed in SeaDataNet can be easily included in ITINERIS metadata catalogue.

Italy is strongly represented in SeaDataNet. 180 institutes are listed in EDMO, 14 of which are SeaDataNet nodes and provide online access to in-situ data and metadata. As the National Oceanographic Data Centre, the OGS coordinates the Italian participation and supports the activities of the Italian data centres.

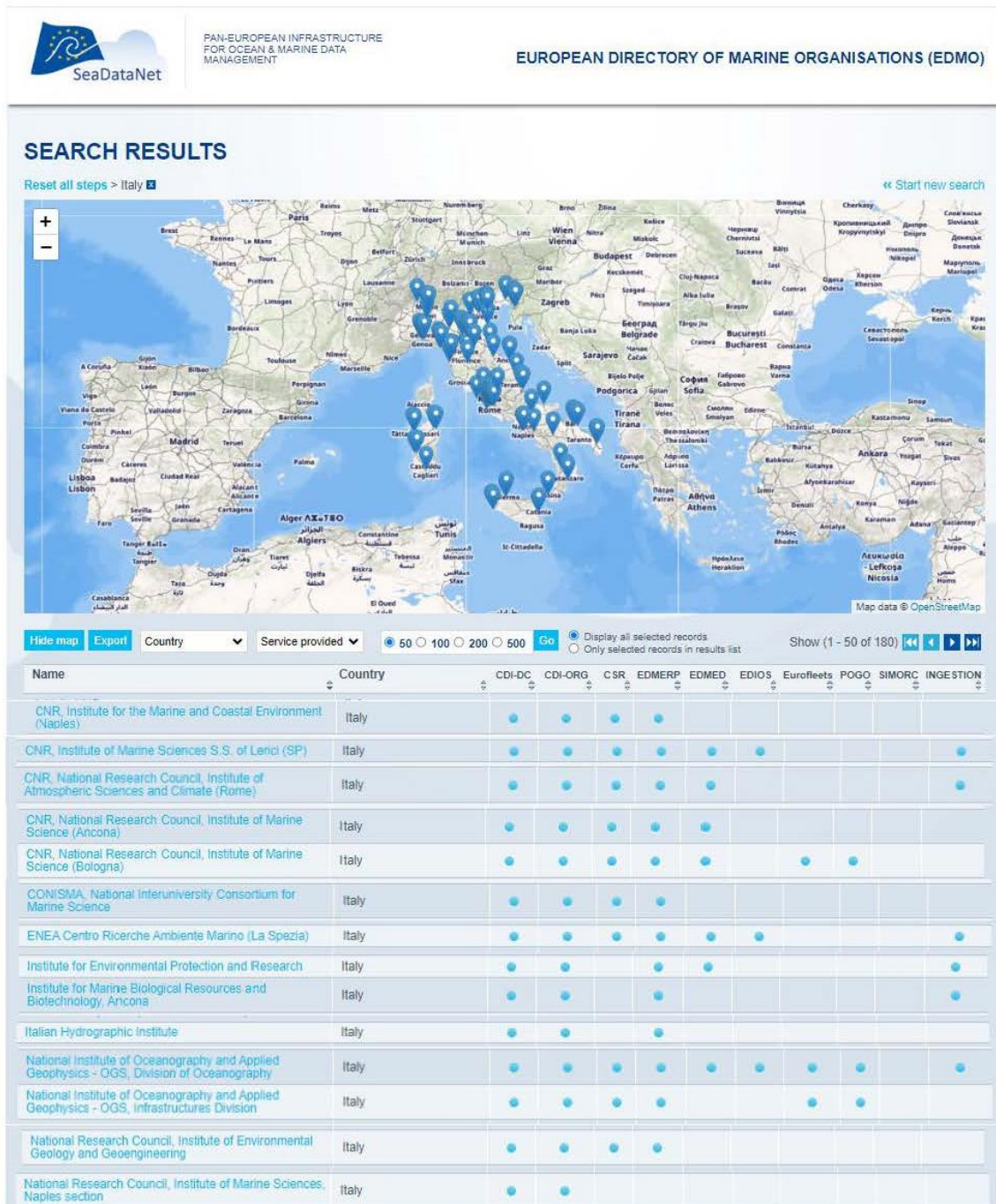


Figure 1 – Italian organisations listed in EDMO as SeaDataNet nodes (https://edmo.seadatanet.org/results/country_italy/00018).

1.7. Applicable and reference documents

This document is applicable to the Italian Integrated Environmental Research Infrastructures System (ITINERIS) Project, and specifically to the development of the IT-IOOS. To ensure compliance with the ITINERIS HUB, it takes into account the specifications defined by WP2.

On this aim the Applicable documents are:

ITINERIS project full proposal.

ITINERIS Deliverable D2.10 – DMP – Data Management Plan.

Reference documents are:

Buck et al., Ocean Data Product Integration Through Innovation-The Next Level of Data Interoperability, *Front. Mar. Sci.*, 2019, Vol. 6, <https://doi.org/10.3389/fmars.2019.00032>.

Piano Nazionale Infrastrutture di Ricerca (PNIR) 2021-2027.

ITINERIS Deliverable D2.5 - Catalogue of data and services, Specification document (First release).

ITINERIS Deliverable D5.1 - Marine Data Centre Management Plan: management needs, service requirements, and governance for the distributed and federated system.

ITINERIS Deliverable D5.2 - Marine Data Store planning with data flows and connection standards for the distributed and federated system interfaced with ITINERIS central hub.

IOC Resolution XXII-6: IOC Oceanographic Data Exchange Policy.

Pecci et al., SeaDataNet, an enhanced ocean data infrastructure giving services to scientists and society, 2020, *IOP Conf. Ser.: Earth Environ. Sci.* 509 012042.

SeaDataNet I3 Pan-European infrastructure for ocean and marine data management - SeaDataNet Data Policy, 2006.

SeaDataNet Guidance for preparation of entries for the European directory of the ocean-observing system (EDIOS), SeaDataNet Deliverable 4.6 B, 29 pp., v2.0, 2013.

1.8. Glossary, Acronyms and Abbreviations

AODN: Australian Ocean Data Network, <https://portal.aodn.org.au/>

API: Application Programming Interface

CDI: Common Data Index

CIOOS: Canadian Integrated Ocean Observing System, <https://cioos.ca/>

CKAN: Comprehensive Knowledge Archive Network

CSR: Cruise Summary Reports

CSW: Catalogue Service for WEB

EDMED: European Directory of Marine Environmental Data

EDMERP: European Directory of Marine Environmental Research Projects

EDIOS: European Directory of Initial Ocean-observing Systems

EDMO: European Directory of Marine Organisations

eLTER-RI: Integrated European Long-Term Ecosystem, critical zone and socio-ecological Research, <https://elter-ri.eu/>

ESFRI: European Strategy Forum on Research Infrastructures

FAIR: Findable, Accessible, Interoperable, and Reusable

GeoSciences: Research infrastructure for the Italian Network of Geological Services, <https://www.isprambiente.gov.it/en/projects/soil-and-territory/geosciences-ir>

EOOS: European Ocean Observing System

EMODnet: European Marine Observation and Data Network

EMSO-ERIC: The European Multidisciplinary Seafloor and water column Observatory, <https://emso.eu/>

EURO-ARGO-ERIC: <https://www.euro-argo.eu/>

EUROFLEETS-RI: An alliance of European marine research infrastructure to meet the evolving needs of the research and industrial communities, <https://www.eurofleets.eu/project-information/>

GOOS: Global Ocean Observing System

ICOS: Integrated Carbon Observation System, <https://www.icos-cp.eu/>

INGV: National Institute of Geophysics and Volcanology, <https://www.ingv.it/>

IOC/UNESCO: Intergovernmental Oceanographic Commission of United Nations Educational, Scientific and Cultural Organization

IODE: International Oceanographic Data and information Exchange

IOOS: Integrated Ocean Observing System (part of NOAA - National Oceanic and Atmospheric Administration), <https://ioos.noaa.gov/>

IT-IOOS: Italian Integrated Ocean Observing System

ISMAR: Institute of Marine Sciences, National Research Council, <https://www.ismar.cnr.it/>

ISP: Institute of Polar Sciences, National Research Council, <https://www.isp.cnr.it/>

ITINERIS: Italian Integrated Environmental Research Infrastructures System, <https://itineris.cnr.it/>

JODC: Japan Oceanographic Data Center, <https://www.jodc.go.jp/jodcweb/>

JERICO-RI: Joint European Research Infrastructure of Coastal Observatories, <https://www.jerico-ri.eu/about/>

N/R Laura Bassi: Research Vessel, owner OGS

LNS: Laboratori Nazionali del Sud, <https://www.lns.infn.it/it/>

MDS: Marine Data Store

MDP: Marine Data Portal

NERC: Natural Environment Research Council

NOAA: National Oceanic and Atmospheric Administration

NODC: National Oceanographic Data Center within the International Oceanographic Data and Information Exchange (IODE) as designated by the Government of the Member State

ODIP: Ocean Data Interoperability Platform

ODIS: Ocean Data and Information System

OGC: Open Geospatial Consortium

OGS: National Institute of Oceanography and Applied Geophysics, <https://www.ogs.it/en>

PU: Production Units of the IT-IOOS Sub-Systems

REST: REpresentational State Transfer

RI: Research Infrastructures

SIOS: Svalbard Integrated Arctic Earth Observing System, <https://sios-svalbard.org/>

SKOS: Simple Knowledge Organization System

SOAP: Simple Object Access Protocol

SPARQL: Simple Protocol and RDF Query Language

SS: Sub-System of the IT-IOOS (the main RIs and two further sub-systems, having the role of accessing and archiving the data)

WFS: Web Feature Service

WMS: Web Map Service

2. KEY FEATURES AND COMPONENTS

2.1. The definition of CKAN dataset

The Marine Data Portal will use CKAN software to publish, share and use ITINERIS data.

CKAN (Comprehensive Knowledge Archive Network) is an open-source data management system for powering data hubs and data portals.

The basic unit in CKAN is called dataset: a dataset object describes a parcel of data that can be a measurement time series from a single sensor for in situ data or homogenous data for a single facility or a network of sensors. A dataset in IT-IOOS can be also a product based on previous collected data.

The dataset contains basically two things (from <https://docs.ckan.org/>):

- Information or “metadata” about the data. For example, the title and publisher, date, what formats it is available in, what license it is released under, etc.
- A number of “resources”, which hold the data itself. CKAN does not mind what format the data is in. CKAN can also store resources by itself but in the IT_IOOS it will store only the metadata catalogue, and the resources will be represented by link to endpoint of the existing subsystems. For example, different resources might contain the data for different years, or they might contain the same data in different formats.

The default metadata schema can be easily customized with the extension ckanext-schema (already tested) and multiple metadata schema can coexist and be managed by a single instance of CKAN.

Using the CSW OGC standard for harvesting remote catalogue multiple instances of CKAN (or other interoperable Catalogue software like Geonetwork) can be organized to compose a federated environment that collect metadata from multiple sources.

The metadata will be mainly copied from each subsystem endpoint: ERDDAP™ data server already support the ISO 19115 format to expose metadata in an interoperable environment compliant with INSPIRE directive requirements.

The IT-IOOS catalogue will rely also in DCAT-AP metadata profile (or the national Italian implementation DCAT-AP-IT) to provide semantic interoperability in compliance with the solutions defined in the former ISA Programme of European Commission and the Interoperable Europe act (<https://eur-lex.europa.eu/eli/reg/2024/903/oj>) recently entered in force.

To consistently set up the IT-IOOS catalogue of metadata a common understanding of the granularity of a CKAN dataset (the main catalogue entry) is necessary.

2.2. IOOS metadata catalogue

The IT-IOOS metadata catalogue lists all ITINERIS datasets. The metadata information for the datasets shall include, in order:

- The metadata of the sub-system,
- The metadata of the facility or observation platform,

- The metadata of the dataset.

2.2.1. Metadata of the Sub-System

<i>ITINERIS Metadata</i>	<i>Description or SeaDataNet reference (EDIOS & vocabs)</i>	<i>Mandatory / Optional</i>
RI name	Name of the Subsystem or Research Infrastructure, free text	M
RI acronym	Free text	M
RI Coordinating organization	Coordinating institute, EDMO reference	M
RI Coordinating person	Principal Investigator (Name/Names, free text)	M
RI Start date	Program Start date	M
RI Geographical coverage	Sea-areas C19 - SEAVOX SALT AND FRESH WATER BODY GAZETTEER	M
RI Abstract	Abstract	O

Table 1 – Metadata related to the Sub-system.

2.2.2. Metadata of the Facility or Observation Platform

<i>ITINERIS Metadata</i>	<i>Description or SeaDataNet reference (EDIOS & vocabs)</i>	<i>Mandatory / Optional</i>
Operating facility name	Platform Name, free text	M
Type of platform (RV, float, offshore structure, fixed benthic node, ...)	Platform class L06 - SEAVOX PLATFORM CATEGORIES	M
Organization owning the facility		M
Organization or institute managing the facility		M
Facility reference person		M

Table 2 – Metadata related to the Facility or Observation Platform.

2.2.3. Metadata of the Dataset

<i>ITINERIS Metadata</i>	<i>Description or SeaDataNet reference (CDI & vocabs)</i>	<i>Mandatory / Optional</i>
Dataset Name or Resource title*	Original title of the dataset, free text	M
Dataset originator or Organization*	EDMO reference	M
Dataset originator, person	Free text	M
Type of device (meteorological packages, sediment traps, Ocean models, ...)	Instrument class L05 - SEADATANET DEVICE CATEGORIES (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L05	M
Instrument (RM Young 32500, Sea-Bird SBE 911 CTD..)	Instrument code L22 - SEAVOX DEVICE CATALOGUE (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L22	O
Discipline	P08 - SeaDataNet Parameter Disciplines (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P08	M
Parameters group	P03 - SeaDataNet Agreed Parameter Groups (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P03	M
Discovery parameter	List of variables, P02 - SEADATANET PARAMETER DISCOVERY VOCABULARY (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P02	M
Temporal extent	Dataset Date and Time or time domain - Start date	M
Data type (NRT/DMT)		M
Spatial extent	Dataset Latitude, longitude or spatial domain - Geographic coverage (bounding box)	M
Coordinate reference system	Coordinate reference system	M
Data licence	Data Access Restriction, L08 - Data Access Restriction (default CC-BY-4.0) (Alt label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L08	M

QC procedures	L12 - SEADATANET PARAMETER QUALITY CONTROL PROCEDURES (Alt label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L12	M
Funding source (e.g. PNRR-ITINERIS)	Project in EDMERP reference	M
Endpoint or Sources*	Endpoint of Data and Resources (Upload or Link)*	M
Resource type*	Web service, Geospatial dataset, Map viewer, ...	O
Original file format		O
Resource Name*	File name	O
Resource Description*		O
DOI		O
Related EDMED dataset		O
Related EDIOS dataset		O
Visibility	Set by the catalogue Administrator (Public/Private)	Internal
State	Set by the catalogue Administrator (active/hidden=deleted)	Internal

Table 3 – Metadata related to the Dataset. *CKAN name for the field.

3. INITIAL IMPLEMENTATION AND MAINTENANCE STRATEGY

3.1. Short-term Strategy

3.1.1. Initial Implementation

In the development of the IT-IOOS metadata catalogue for data and facilities, the following steps are planned:

1. A collection phase of the information from different sources. This concerns also the review of existing metadata delivered in SeaDataNet directories (deadline: December 2024)
2. A compilation phase by all the IT-IOOS Sub-systems for the new IT-IOOS catalogue entries. This is done standardizing metadata from all participating RIs towards the agreed IT-IOOS metadata catalogue format (deadline January 2025)
3. A reviewing phase with the integration of all metadata in CKAN (deadline February 2025)

The IT-IOOS metadata catalogue is functional for the Backend (Orchestrator) component of the Marine Data Portal, that will manage its content to enable users to search for data, products and services following a wide range of search criteria. The IT-IOOS metadata catalogue provides information for the Website component. Infrastructures that do not yet organize and structure data will find themselves with a form for organizing and publishing metadata.

Everyone else will see their forms mapped to the IT-IOOS one (orchestrator's job)

3.1.2. Operational Support

- Setting up an internal support team with representatives of the IT-IOOS systems for ongoing metadata updates and user assistance.
- Developing training programs for RI personnel on metadata standards and submission procedures.

3.2. Long-term Strategy

The IT-IOOS metadata catalogue is based on standards that are supported and fostered by national and international scientific communities. The adoption and further development of open-source approaches and tools enables the implementation of new data workflows. The modern paradigm is moving beyond data portals towards service-based architectures that combine data provenance, persistence, and security (both physical and online). In this way, metadata, data and facilities aims at being FAIR and delivering via web services, using the concept of federated services to facilitate the interoperability of data and bring together the different information from the IT-IOOS subsystems.

Additionally, the IT-IOOS Hub will be moving beyond a classic data portal and integrating automated and customizable search algorithms, advancing the FAIR principles and embracing a more dynamic, intelligent approach to data discovery and reuse. In the context of marine data, which is often large-scale and interdisciplinary, the integration of such intelligent systems will enable more nuanced, meaningful use of data. This approach will transform how researchers, policymakers, and other stakeholders interact with and benefit from ocean data.

The endorsement of such architecture allows the ITINERIS community to develop services that meet its specific national needs while maintaining data interoperability with the international framework.

The adoption of standards, platforms and data services supported by scientific communities guarantees the sustainability of infrastructure and services and thus the maintenance and improvement of the IT-IOOS metadata catalogue. This requires collaboration with national and international marine data initiatives and continuous alignment with global standards and practises.

Through active engagement of the communities that depend on these tools, the burden of support, regular adaptation to technological advances and regular updating of content and documentation can be shared to reduce the burden on a single institution and ensure the sustainability of the metadata catalogue in the long term.

3.3. Alignment with the ITINERIS Catalogue of data and services

The ITINERIS Catalogue of Resources (namely RIs, Providers, Services, Dataset, Research products, Training resources and VREs) is one of the main components of the ITINERIS HUB, realized to collect and organize all the knowledge, data, analytics tools, and services, and to serve as a common, comprehensive, and discoverable online registry of the resources offered by the ITINERIS project. The adoption of standardized metadata schemas and protocols helps to translate and align existing metadata formats, ensuring seamless integration and accessibility of all resources within the unified catalogue. To this end, metadata schemas/profiles have been identified for each seven resources.

The collection of metadata should be possible via metadata services that implement Catalogue Service for WEB (CSW), a standard of the Open Geospatial Consortium (OGC). The dataset is catalogued in ISO/TS 19139, a metadata schema based on the international standard for the representation of geographic metadata (gmd), which provides the XML implementation schema derived from ISO 19115.

The OGC catalogue interface standards specify the interfaces, bindings and framework for defining application profiles required for publishing and accessing digital catalogues of geospatial metadata, services and related resource information. Metadata acts as generalised properties that can be queried

and returned via catalogue services to evaluate resources and, in many cases, to call up or retrieve the referenced resource.

In the case of the marine domain, a thematic catalogue has been defined and will serve as collector of metadata from the 11 marine RIs. The IT-IOOS metadata catalogue for data and facilities is based on the best practices at international level, as e.g. SeaDataNet catalogues for data and metadata that conform to ISO standards. For example, the IT-IOOS metadata catalogue includes the EDMO (European Directory of Marine Organisations) code, which contains up-to-date addresses and activity profiles of research institutes, data centres, monitoring agencies, government and private organisations involved in one way or another in oceanographic and marine research, data and information management and/or data collection. Currently, more than 4,000 organisations are listed and described in EDMO.

EDMO offers two interfaces, a normal HTML interface for "human users" and a SPARQL endpoint for machine applications (<https://www.seadatanet.org/Metadata/EDMO-Organisations>). These information can be imported in standard format in IT-IOOS and can be federated within other frameworks.

In addition, the IT-IOOS metadata catalogue for data and facilities adopts the SeaDataNet controlled vocabularies, which are checked and published using standardised SKOS APIs (REST, SOAP and SPARQL).

The list of metadata proposed by the IT-IOOS catalogue to describe the datasets is based on the metadata proposed by the SeaDataNet infrastructure. This will facilitate the integration of IT-IOOS in ITINERIS-HUB both in the catalogue description and in the data harvesting.

3.4. Alignment with SeaDataNet Catalogues

The alignment of the IT-IOOS catalogue with the SeaDataNet catalogue will ensure the integration and visibility of ITINERIS data in the wider international landscape. In fact, SeaDataNet infrastructure is adopted by the European Marine Observation and Data Network (EMODnet), the long-term initiative of the European Commission to provides easy and free access to in-situ marine data, metadata and data products in Europe.

With this purpose, the list of metadata included in the IT-IOOS catalogue to describe the marine datasets and facilities in the marine domain was created by selecting the information available in the SeaDataNet catalogues. This enabled the reuse of this information for the compilation and insertion of the following catalogues within the SeaDataNet infrastructure:

- EDMO for the organisations,
- EDIOS (PROGRAM for the EDIOS SERIES, the EDIOS PLATFORM) for the observing systems,
- EDMED for the datasets.

To facilitate the collection and sharing of information, it will be possible to extract the metadata in tabular form from the marine HUB to use the Mikado software (<https://www.seadatanet.org/Software/MIKADO>) to compile and standardise the information.

4. ANNEX - METADATA CATALOGUE FORM

Subsystem or Research Infrastructure		
ITINERIS Metadata	Description or <u>SeaDataNet</u> reference (EDIOS & vocabs)	Mandatory / Optional
RI name	Name of the Subsystem or Research Infrastructure	M
<u>RI acronym</u>	Free text	M
RI Coordinating institute	Coordinating institute, EDMO reference	M
RI Coordinating person	Principal Investigator (Name/Names, free text)	M
RI Start date	Program Start date	M
RI Geographical coverage	Sea-areas C19 - SEAVOX SALT AND FRESH WATER BODY GAZETTEER	M
RI Abstract	Abstract	O
Facility or Observing platform:		
Operating facility name	Platform Name, free text	M
Type of platform (RV, float, offshore structure, fixed benthic node, ...)	Platform class L06 - SEAVOX PLATFORM CATEGORIES	M
Organization owning the facility		M
Organization or institute managing the facility		M
Facility reference person		M
Dataset		
Dataset Name or Resource title*	Original title of the dataset, free text	M
Dataset originator or Organization*	EDMO reference	M
Dataset originator, person	Free text	M
Type of device (meteorological packages, sediment traps, Ocean models, ...)	Instrument class L05 - SEADATANET DEVICE CATEGORIES (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L05	M
Instrument (RM Young 32500, Sea-Bird SBE 911 CTD ...)	Instrument code L22 - SEAVOX DEVICE CATALOGUE (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L22	O
Discipline	P08 - <u>SeaDataNet</u> Parameter Disciplines (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P08	M
Parameters group	P03 - <u>SeaDataNet</u> Agreed Parameter Groups (Preferred label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P03	M
Discovery parameter	List of variables, P02 - SEADATANET PARAMETER DISCOVERY VOCABULARY (Preferred label)	M

	https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=P02	
Temporal extent	Dataset Date and Time or time domain - Start date	M
Data type (NRT/DMT)		M
Spatial extent	Dataset Latitude, longitude or spatial domain - Geographic coverage (bounding box)	M
Coordinate reference system	Coordinate reference system	M
Data licence	Data Access Restriction, L08 - Data Access Restriction (default CC-BY-4.0) (Alt label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L08	M
QC procedures	L12 - SEADATANET PARAMETER QUALITY CONTROL PROCEDURES (Alt label) https://vocab.seadatanet.org/v_bodc_vocab_v2/search.asp?lib=L12	M
Funding source (e.g. PNRR-ITINERIS)	Project in EDMERP reference	M
Endpoint or Sources*	Endpoint of Data and Resources (Upload or Link)*	M
Resource type*	Web service, Geospatial dataset, Map viewer, ...	O
Resource Name*	File name	O
Original file format		O
Resource Description*		O
DOI		O
Related EDMED dataset		O
Related EDIOS dataset		O
Visibility	Set by the catalogue Administrator (Public/Private)	Internal
State	Set by the catalogue Administrator (active/hidden=deleted)	Internal