

Project Title	FAIR EArth Sciences & Environment services
Project Acronym	FAIR-EASE
Grant Agreement No.	1010587
Start Date of Project	1/09/2022
Duration of Project	36 Months
Project Website	fairease.eu

D2.1 Environmental Data Infrastructures: Services Description Report

Work Package	WP2, Discovery, Access and FAIR Data services
Lead Author (Org)	Tjerk Krijger (MARIS)
Contributing Author(s) (Org)	Enrico Boldrini (CNR-IIA), Roberto Roncella (CNR-IIA), Fabrizio Papeschi (CNR-IIA), Alexandra Kokkinaki (NOC-BODC), Gwen Moncoiffe (NOC-BODC), Véronique Chaffard (Theia - IRD), Peter Thijsse (MARIS)
Due Date	31.03.2023
Date	11.04.2023
Version	V1.0

Dissemination Level

- PU: Public
- PP: Restricted to other programme participants (including the Commission)
- RE: Restricted to a group specified by the consortium (including the Commission)
- CO: Confidential, only for members of the consortium (including the Commission)

Versioning and contribution history

Version	Date	Author	Notes
0.1	24.02.2023	MARIS	First version
0.2	10.03.2023	MARIS, CNR-IIA, NOC-BODC, IRD	Draft version
0.3	20.03.2023	MARIS, CNR-IIA, NOC-BODC, IRD	Final Draft
0.4	24.03.2023	MARIS, CNR-IIA, NOC-BODC, IRD, VLIZ, CNRS	Final Version
0.5	30.03.2023	MARIS, CNR-IIA, NOC-BODC, IRD, VLIZ, CNRS, NEOVIA	Final edition for submission
1.0	11.04.2023	Neovia	Final version approved by the Technical Board

Disclaimer

This document contains information which is proprietary to the FAIR-EASE Consortium. Neither this document nor the information contained herein shall be used, duplicated or communicated by any means to a third party, in whole or parts, except with the prior consent of the FAIR-EASE Consortium.

Table of contents

Executive Summary.....	6
1 Introduction.....	7
1.1 Overall concept.....	7
1.1.1 Metadata brokerage.....	8
1.1.2 Data brokerage.....	9
1.1.3 Smart solutions.....	9
1.2 Next steps.....	9
1.3 Prioritisation of Data Infrastructures.....	10
2 Priority (Data) Infrastructures.....	11
2.1 Copernicus Marine Service (CMEMS).....	11
2.2 EMODnet Chemistry, Physics, Bathymetry, Biology.....	15
2.3 Euro-Argo.....	19
2.4 SeaDataNet.....	20
2.4.1 SeaDataNet Common Data Index (CDI).....	20
2.4.2 SeaDataNet Data Products & Sextant web catalogue.....	23
2.5 GEBCO.....	25
2.6 NASA Earth Observation Data.....	27
2.7 Copernicus Land Monitoring Service.....	29
2.8 European Space Agency.....	32
2.8.1 Copernicus Open Access Hub.....	32
2.8.2 EUMETSAT.....	35
2.9 Climate Data Store.....	38

2.10 NOAA.....	40
2.10.1 Physical Sciences Laboratory	40
2.10.2 National Centers for Environmental Information	41
2.11 European Soil Data Centre.....	43
2.12 Joint Research Centre Data Catalogue	44
2.13 European Environment Agency	47
2.14 National Center for Biotechnology Information (NCBI)	51
2.15 Global Biodiversity Information Facility (GBIF).....	53
2.16 Ocean Biodiversity Information System (OBIS)	55
2.17 European Nucleotide Archive (ENA).....	57
2.18 World Register of Marine Species	59
2.19 European Marine Omics Biodiversity Observation Network	61
3 Other (Data) Infrastructures.....	62
3.1 European Gliding Observatories (EGO)	62
3.2 Marine Mammals Exploring the Oceans Pole to Pole (MEOP).....	62
3.3 Glodap (Global Ocean Data Analysis Project).....	62
3.4 Global Runoff Data Centre (GRDC)	63
3.5 Global Historical Climatology Network - Daily (GHCN-Daily)	63
3.6 Marine Regions gazetteer.....	63
3.7 BiG-FAM database	64
3.8 AERONET Aerosol Network.....	64
4 Conclusions.....	65

Table of figures

FIGURE 1 EXAMPLE OF HOW AN ARGO FLOAT PROFILE FUNCTIONS.....	19
--	----

TERMINOLOGY

Terminology/Acronym	Description
DoA	Description of Action
API	Application Programming Interface
BC	Blue-Cloud
BDI	Blue Data Infrastructures
BGCs	Biosynthetic Gene Clusters
BODC	British Oceanographic Data Centre
CDI	Common Data Index
CDS	Climate Data Store
CSW	Catalogue Service for the Web
DAB	Discovery and Access Broker
DDBJ	DNA Data Bank of Japan
DTM	Digital Terrain Model
EAL	Earth Analytical Lab
EC	European Commission
EEA	European Environment Agency
EGO	European Gliding Observatories
EMBL	European Molecular Biology Laboratory
EML	Ecological Metadata Language
EMODnet	European Marine Observation and Data network
ENA	European Nucleotide Archive
EOSC	European Open Science Cloud
ESDAC	European Soil Data Centre
ESDS	Earth Science Data Systems
FAIR	Findable; Accessible; Interoperable; Reusable
FE	FAIR-EASE
FDBAQ	Flexible Dynamic Block Adaptive Quantization
GA	Grant Agreement to the project
GCF	Biosynthetic Gene Cluster Family
GBIF	Global Biodiversity Information Facility
GEBCO	General Bathymetric Chart of the Oceans
GHCN - Daily	Global Historical Climatology Network - Daily
GLODAP	Global Ocean Data Analysis Project
GRD	Ground Range Detected
GRDC	Global Runoff Data Centre
GUI	Graphical User Interface
IDDAS	Interdisciplinary Data Discovery and Access Service
INSDC	International Nucleotide Sequence Database Collaboration
IPT	Integrated Publishing Toolkit

Terminology/Acronym	Description
JRC	Joint Research Centre
KPI	Key Performance Indicator
LRM	Low-Resolution Mode
MEOP	Marine Mammals Exploring the Oceans Pole to Pole
NCBI	National Center for Biotechnology Information
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
OBIS	Ocean Biodiversity Information System
OLCI	Ocean and Land Color Instrument
OSW	Ocean Swell spectra
OWI	Ocean Wind field
OWL	Ontology Web Language
PSL	Physical Sciences Laboratory
RI	Research Infrastructure
RVL	Surface Radial Velocity
SAR	Synthetic-Aperture Radar
SLC	Single Look Complex
SLSTR	Sea and Land Surface Temperature Radiometer
SOAP	Simple Object Access Protocol
UC	Use Case
VRE	Virtual Research Environment
WoRMS	World Register of Marine Species
WMS	Web Mapping Service

Executive Summary

For many societal and scientific challenges of the present, access to multi-disciplinary data resources and services is key. The FAIR-EASE (FE) project presents a technical framework that tries to cross boundaries of domain-specific (meta)data standards that would allow for developing overarching services where scientists from different domains can collaborate. This project focuses on three different multi-disciplinary Use Cases (UCs) that specify the needs and requirements for their research.

One of the components of the FE technical framework will be the FE Interdisciplinary Data Discovery and Access Service (IDDAS) that will provide users with an easy and FAIR service for discovery and access to multidisciplinary and aggregated data sets. These data resources are managed and provided by a range of existing environmental data infrastructures that operate in different domains such as Biodiversity, Atmosphere, Solid Earth and Ocean. The data infrastructures that are included in this project are based on the needs and requirements set by the UCs.

The federation of data resources provided by the data infrastructures should facilitate the sharing of multi-disciplinary datasets with users and allow for interaction with the FE Data Lake (WP4) and Earth Analytic Lab (EAL - WP3) that will grant users access to analytical and visualisation services and applications.

For each of the data infrastructures chosen by the UCs to include in this project, this Deliverable D2.1 includes an analysis of their current data discovery and access mechanisms, as well as a first analysis of metadata models and semantics used. The content of this report is important for a common understanding of the data infrastructure landscape, and will e.g. provide the URLs that the IDDAS metadata harvester (based on [GEODAB](#)) will use.

1 Introduction

Many societal and scientific challenges of the present are intertwined, meaning that aspects from different domains are included. It is therefore important to kick-start multidisciplinary projects that allow scientists from different domains to collaborate. One of the prerequisites for these types of projects is to cross boundaries of domain-specific metadata standards that would allow for developing overarching services. The FAIR-EASE (FE) project is such a research initiative that tries to achieve this by creating a technical framework with the aim to meet the needs set by three different multi-disciplinary Use Cases (UCs).

One of the components of the FE technical framework is the **Interdisciplinary Data Discovery and Access Service (IDDAS)**, which will provide users with an easy and FAIR service for discovery and access to multidisciplinary and aggregated data sets (in situ datasets, satellite datasets, omics experiments and model outputs) as managed and provided by a range of existing environmental data infrastructures. The federation of data resources should facilitate the sharing of multi-disciplinary datasets with users, with the FE data lake, and allow for interaction with the Earth Analytic Lab (EAL) that will grant users access to analytical and visualisation services and applications.

The FE IDDAS is based on the Blue-Cloud (BC) DDAS that presents the latest evolution of a federation of Blue Data Infrastructures (BDIs) in the European Ocean domain that allows for discovery and access to a set of BDIs, managed by Research Infrastructures (RIs) and leading European Union (EU) initiatives: EMODnet, EurOBIS, SeaDataNet, ICOS, SOCAT, ELIXIR-ENA, EcoTaxa, and Euro-Argo.

In the FE project one of the major challenges will be to integrate data infrastructures from a wide range of different domains into the FE IDDAS. Furthermore, the FAIR-EASE Interdisciplinary Data Discovery and Access Service will use the metadata from these data infrastructures as it is offered. Indeed, improving their richness is not in the scope of the FAIR-EASE project as many (if not “most”) of the data infrastructures are outside of the project. FE will have to work with the services, metadata and dataformats as provided, and can only provide feedback for improvement.

For each of the data infrastructures that are included in the UCs, it is important to conduct a thorough analysis of their current data discovery and access mechanisms. This Deliverable D2.1 has selected the most important data infrastructures for the UCs and performs an analysis on these.

1.1 Overall concept

As mentioned above, the technical architecture of the FE IDDAS will be based on the BC DDAS for which the overall concept is taken from BC [D2.1](#) and modified here to fit the requirements of the FE IDDAS.

Part of the concept is to rely on a fully machine-to-machine interaction between the FE IDDAS and each of the data infrastructures. Another part of the concept is a two-step approach to discovery from data collections (e.g. an Argo float) to data granules (e.g. one cycle of such a float). For this two-step approach, the FE IDDAS implements brokerage services at two levels; 1) collection level and 2) granule level, which will provide the user/machine with discovery to:

1. Collection level: The data collections at metadata level, using a to-be-determined ISO-19115-1:2014 based FE metadata profile as target. Note that ISO 19115 contains more than 400 metadata elements that include mandatory, optional or conditional elements and that the FE metadata profile will identify the most important ones for the FE context and add elements if necessary. For level 1, the focus lies on identifying the data at an aggregated collection level, using search options like, free search, geographic search and temporal filters.

2. Granule level: The individual datasets as they are published by the federated data infrastructures; their metadata description is retrieved through heterogeneous web services where available, depending on the data provider (e.g., OGC-CSW, ERDDAP, APIs). Here the focus lies on drilling down (facet search) within identified collections to get more specific data records at granule level leveraging as well additional search criteria which are specific to each data infrastructure. This second step to granule level services could be very challenging and it will have to be investigated for with data infrastructures this is possible.

The access links in the FE metadata catalogue will act as pointers to the FE data brokerage service that will interact with the data access API's of the data infrastructures. This allows users to discover data collections at the FE metadata catalogue, with possibility for further access by downloading their selected collection, all through one common FE IDDAS interface. Below the metadata and data brokerage functionalities are described.

1.1.1 Metadata brokerage

The FE IDDAS will harvest metadata from the data infrastructures using the GEODAB metadata brokerage mechanism. This is done at collection level from each of the data infrastructures, using their indicated web services or APIs. The DAB service then transforms the harvested files (encoded according to different standards) from the data infrastructures into a ISO19115 based FE collection metadata profile, which is then published by the DAB service by means of a CSW service supporting an harmonised XML encoding for machine-to-machine interactions, as well as by GUI for human users.

The FE metadata model will mandatory include a limited number of metadata tags, to be completed by mappings with each of the metadata services of the data infrastructures. The common metadata model will focus on describing data collections by means of fields for What, When, Where, Who, and additional fields for querying more detailed metadata at each source and for facilitating access by means of downloading the data collections from the source infrastructures.

The mapping of the metadata models of the data infrastructures to the FE metadata model, will be supported by the use of vocabularies. This semantic brokerage is required as data infrastructures from different domains use different vocabularies for metadata fields

Use of common vocabularies in all metadatabases and data formats is an important prerequisite towards consistency, interoperability, and FAIRness. Common vocabularies consist of lists of standardised terms that cover a broad spectrum of disciplines of relevance to the UCs involved in the FE project. Using standardised sets of terms solves the problem of ambiguities associated with data markup and also enables records to be interpreted by computers. This opens up data sets to a whole world of possibilities for computer aided manipulation, distribution and long-term reuse. Therefore, for the FE IDDAS it will be explored if common vocabularies can be adopted and deployed in the FE IDDAS, and where possible setting up semantic ([smart](#)) mappings between sources and targets. A very good candidate for targets, and storage of new vocabularies, and support for mappings, is the set of common vocabularies which have been initiated and are maintained by SeaDataNet (and led by BODC) for the marine and oceanographic data community.

The SeaDataNet common vocabularies contains multiple lists of standardised terms that cover a broad spectrum of disciplines of relevance to the oceanographic and wider community. It follows the W3C SKOS specification for encoding data dictionaries and taxonomies served. The SeaDataNet vocabulary services are technically managed and hosted by the British Oceanographic Data Centre (BODC) at the [NERC Vocabulary Server \(NVS2.0\)](#).

The SeaDataNet controlled vocabularies are made available as web services and RDF linked data for machines and by means of client interfaces for end-users. The client interfaces provide end-users options for searching, browsing and CSV-format export of selected entries. The machine interfaces are provided via a SOAP API and as SPARQL endpoints.

The adoption of SeaDataNet vocabularies will be considered where mappings to vocabularies from other disciplines are possible and not too extensive in order to contribute to harmonisation of metadata and data, improving the interoperability. The SDN vocabularies and other vocabs will also support the metadata of the input and output data(products) from the planned Virtual labs/Pilots resulting from work in the pilots, that also will be published again as part of the FAIR-EASE IDDAS.

1.1.2 Data brokerage

For the data access part of the FE IDDAS, there will be a data brokerage service developed, for which the data delivery mechanisms of the data infrastructures must be analysed. It will have to be determined what type of data access mechanisms the FE IDDAS will entail, but the planned approach is to adopt the existing API's of the data infrastructures and connect them to the central part of the FE data brokerage service. The data brokerage service will specifically be used to support access/synchronisation of dedicated data lakes (with e.g. subsetting services and visualisation) to support the work in the EAL.

It should be noted that the retrieved data sets from the data infrastructures will not be harmonised, but will be provided in the original data formats.

1.1.3 Smart solutions

The FE IDDAS is firstly considered as a direct interfacing to the data resources of the data infrastructures. But in some cases, a 'smart' federation might be preferred in cases where for example the data infrastructure provides local applications for sub-setting/filtering images that are useful for the end user. Here, the data resources could first be pre-processed at a data infrastructure to for example reduce the volume or provide added-value before transferring the resource to the FE IDDAS. These cases require 'smart' federation and will have to be identified.

1.2 Next steps

In the following months a more thorough analysis will be performed to detail technical specifications of the FE IDDAS. It will be investigated to what degree it is possible to map the metadata of the data infrastructures to the to-be-determined FE metadata model. A common metadata broker will be proposed that will feed the FE collection-level catalogue service component. An interdisciplinary data brokerage service where existing web services by which data sets can be identified will be also set up. The data collection discovery and data discovery and brokerage service components will be integrated into the FE IDDAS.

These next steps, including the technical specification and workplan for the implementation and deployment of the FE IDDAS will be documented in Deliverable D2.2 at M12. Thereafter, in the months till M18, the actual developments and deployment will take place for establishing a first release of the FE IDDAS.

1.3 Prioritisation of Data Infrastructures

At the start of this project an inventorisation of the relevant data infrastructures has been performed among the UCs. This has led to a document containing a list of data infrastructures crossing multiple domains such as Atmosphere, Ocean, Land, Biodiversity, Marine Omics and Meteorology. The document summarises the data resources that are expected to be employed to implement the demonstrators planned within each Pilot. The document is released as Annex A of [Deliverable D5.1](#) and it is also expected to evolve during the course of the project to represent the reference resources of interest in FAIR-EASE and will also be included in the project DMP.

The link to the mapping table that will be maintained during the course of the whole project is available at:

- https://docs.google.com/spreadsheets/d/18uRj4MFaCDJPcbcuscy1KdTG95h_imhj/edit#gid=1818952351

From this document a selection of priority (data) infrastructures is made that will be analysed in Chapter 4. The other (data) infrastructures are mentioned with a short description in Chapter 5, but are not analysed in detail.

2 Priority (Data) Infrastructures

2.1 Copernicus Marine Service (CMEMS)

2.1.1 Data Discovery and Access Component

The Copernicus Marine Service includes the Copernicus Marine Data Store that gives access to a large number of ocean data products.

2.1.1.1 Name & Web Address

Copernicus Marine Service: <https://marine.copernicus.eu/>

2.1.1.2 Types and number of data sets and/or data products

There are currently 275 ocean data products available via the Marine Data Store that include hindcast, current and forecast data based on satellite observations, numerical models or in-situ observations. There are 16 ocean variables for which these products are available, ranging from nutrients to temperature to the carbonate system spanning the global ocean or local seas.

2.1.1.3 Discovery and Access mechanism and functionality

For discovery and access there is the Copernicus Marine Data Store Catalogue:

- <https://data.marine.copernicus.eu/products>

This catalogue contains a GUI where the user can use free search to find the relevant data product and apply filters based on variables, areas and sources.

After finding the right data product there are multiple data access options available from the GUI that are documented [here](#):

- MOTU
- OPeNDAP
- ERDDAP
- FTP
- WMS (visualisation)

Credentials are in general needed for downloading data. They can be obtained through a simple registration process.

The data product can also be visualised on the CMEMS map viewer from where the data can also be downloaded:

- <https://data.marine.copernicus.eu/viewer/expert>

2.1.2 Semantic Interoperability

2.1.2.1 Metadata format(s)

Each data product has a corresponding metadata file in XML using the metadata encoding standard ISO 19139.

The ISO 19139 MyOcean profile v.0.2 is referenced to be used by the service to precisely indicate the available subset of ISO 19115 metadata elements. However, a complete description of this profile couldn't be found, however, from the online documentation, at least the following metadata elements should be present: Identifier, date, title, type, subject(s) (keywords and variable names), abstract with

DOI, description, rights, language, source, format (NETCDF-4), Bounding Box values, endpoints URL to use FTP/MOTU-SUBS/WMS:GetCapabilities protocols for each dataset of the "identified" product.

2.1.2.2 Data format(s)

The data products are downloaded in NetCDF format and follow CF-1.4 or CF-1.6 [convention](#). The Climate and Forecast Standard names conventions define metadata of what the data in each variable represents, and the spatial and temporal properties of the data.

2.1.2.3 Use of controlled vocabularies - which, where, how

Use in metadata

Controlled vocabulary	Vocabulary URL	Example terms	Comment
GEMET - INSPIRE themes, version 1.0	P22	Oceanographic geographical features	thematic classification
My Ocean ontology "Temporal scale" concepts		multi-year	temporal scale
My Ocean ontology "Discipline" concepts		In-situ-observation, satellite-observation	discipline
My Ocean ontology "Area of benefit" concepts		marine-resources	Area of benefit
My Ocean ontology "Reference Geographical Areas" concepts		global-ocean	reference geographical area (gazetteer)
My Ocean ontology "Processing level" concepts		N/A	processing level
My Ocean ontology "Model assimilation" concepts		Not Applicable	model assimilation
My Ocean ontology "Feature type" concepts		Po int series	feature type
Climate and Forecast Standard Names	P07	integral_of_sea_water_temperature_wrt_depth_expressed_as_heat_content	parameter
Copernicus marine in situ	P09	MEDATLAS Parameter Usage Vocabulary	parameter

Use in data

The Climate and Forecast Standard Names are used to describe the variables in the datasets, the related [P07](#) list can be found on the Nerc Vocabulary Service and the corresponding units follow the [P06](#) list.

4.1.3 Available web services for metadata access

The Catalogue Service for the Web (CSW) is used to discover the Copernicus Marine Catalogue and its related products, datasets and web services from this endpoint:

- [HTTPs - CSW API](#)

Discovery metadata endpoint: Custom JSON API

The CMEMS catalogue uses a public REST API that can be leveraged by software clients for machine to machine interactions. The base endpoint of the API is:

- <https://cmems-be.lobelia.earth/api>

Two paths are available in the API, depending on the managed resource type. They are:

- **dataset** path: used to discover resources of type “dataset”
- **metadata** path: to obtain “metadata” resources describing a “dataset” resource in detail.

The **dataset** path can be used to query the catalogue and obtain matching dataset descriptions. Possible query parameters include free text, time range, facets (e.g. depth, universe, main variables, area, indicators & trends, feature type, temporal resolution, source, processing level, EU directive, blue markets, originating centre), pagination.

The results are JSON encoded and include the metadata identifier to be used with the metadata operation. Example result:

```
"GLOBAL_OMI_OHC_area_averaged_anomalies_0_2000": {
  "title": "Global Ocean Heat Content (0-2000m) time series and trend from Reanalysis & Multi-Observations Reprocessing",
  "thumbnailUrl": "https://catalogue.marine.copernicus.eu/documents/IMG/GLOBAL_OMI_OHC_area_averaged_anomalies_0_2000.png",
  "omiFigureUrl": "https://catalogue.marine.copernicus.eu/documents/IMG/GLOBAL_OMI_OHC_area_averaged_anomalies-hq.png",
  "sources": [
    "In-situ observations",
    "Numerical models",
    "Satellite observations"
  ],
  "areas": [
    "Global Ocean"
  ],
  "tempResolutions": [
    "Yearly"
  ],
  "mainVariables": [
    "Temperature"
  ],
  "numLayers": 1
},
```

The **metadata** path is used to retrieve metadata record describing the dataset according to ISO 19115 data model (XML encoded using ISO 19139 schema).

Metadata operation template:

- <https://cmems-be.lobelia.earth/api/metadata/{metadata-identifier}>

Example metadata operation request:

- https://cmems-be.lobelia.earth/api/metadata/GLOBAL_OMI_OHC_area_averaged_anomalies_0_2000

Discovery metadata endpoint: CSW

Alternatively, an OGC CSW endpoint (AP ISO profile) is available at the endpoint:

- <https://cmems-catalog-ro.cls.fr/geonetwork/srv/eng/csw-MYOCEAN-CORE-PRODUCTS?>

The OGC CSW specification is described at:

- <https://www.ogc.org/standard/cat/>

A CSW catalogue can be queried or harvested using its GetRecords operation.

The GetCapabilities operation shows the catalogue capabilities (e.g. supported metadata standards, query constraints, etc.):

- <https://cmems-catalog-ro.cls.fr/geonetwork/srv/eng/csw-MYOCEAN-CORE-PRODUCTS?SERVICE=CSW&REQUEST=GetCapabilities>

Support for generic ISO 19139 metadata documents is present. The GetRecordById operation is used to retrieve the description of a specific metadata record. Example request:

- <https://cmems-catalog-ro.cls.fr/geonetwork/srv/eng/csw-MYOCEAN-CORE-PRODUCTS?service=CSW&request=GetRecordById&version=2.0.2&outputSchema=http://www.isotc211.org/2005/gmd&ElementSetName=full&id=2ed3a04a-7dca-4eb0-b0e6-22365a3273a0>

4.1.4 Available web services for data access

4.1.4.1 Data access policy - if yes, which and how deployed

Open and free access to the data, registration is required.

4.1.4.2 Any web services and API's - URLs, function, how to operate

The Copernicus Marine Service provides multiple APIs that are INSPIRE compliant and OGC standardised. More information about the APIs can be found from the CMEMS documentation and in the sections below:

- [HTTPs - MOTU API](#)
- [HTTPs - WMS API](#)
- [HTTPs - WCS API](#)
- [HTTPs - ERDDAP API](#)
- [HTTPs - OPeNDAP API](#)
- [FTP - File Transfer Protocol](#)

Access data endpoint: Custom API

API are as well available for dataset download. A sample API request can be obtained from the CMEMS map viewer clicking on the button “Show API request”. Credentials are needed as well in this case.

2.2 EMODnet Chemistry, Physics, Bathymetry, Biology

2.2.1 Data Discovery and Access Component

The European Marine Observation and Data network (EMODnet) connects organisations supported by the EU's Integrated Maritime Policy who work together to observe the sea, process the data according to international standards (relevant to topic, e.g. EEA, GEBCO, etc) and make that information freely available as interoperable data layers and data products.

2.2.1.1 Name & Web Address

European Marine Observation and Data network: <https://emodnet.ec.europa.eu/en>

2.2.1.2 Types and number of data sets and/or data products

EMODnet provides easy and free access to marine data, metadata and data products and services spanning seven discipline-based themes: bathymetry, biology, chemistry, geology, human activities, physics and seabed habitats. Each theme is dealt with by a partnership of organisations that possess the expertise necessary to standardise the presentation of data and create data products.

In this analysis the themes Bathymetry, Chemistry, Physics, and Biology are highlighted.

Bathymetry

EMODnet Bathymetry provides a service for viewing and downloading the best available harmonised Digital Terrain Model (DTM) for the European sea regions, together with a range of other bathymetric data, products, and services. The DTM is generated and regularly updated by the EMODnet Bathymetry partnership using an increasing number of bathymetric data sets. The current 2020 release of the EMODnet Digital Terrain Model has a grid resolution of 1/16 * 1/16 arc minutes.

Chemistry

EMODnet Chemistry provides access to marine chemical data, standardised harmonised validated data collections and reliable data products. EMODnet Chemistry focuses on four main themes:

- [Eutrophication](#);
- [Ocean acidification](#);
- [Contaminants](#);
- [Marine litter](#).

Observational datasets include seawater quality, biota contamination, sediment quality and potential pollution. The measurement data cover 13 [groups](#) of chemical variables within all European sea regions.

Physics

EMODnet-Physics provides access to in situ ocean physics data, including temperature, salinity and currents profiles, sea level trends, wave height and period, wind speed and direction, water turbidity (light attenuation), underwater noise, river flow, and sea-ice coverage.

Measurement types include fixed platforms (moorings, tide gauges, HF radars, etc.), moving platforms (ARGO, Lagrangian buoys, ferryboxes, etc.) and repeated observations (CTDs, etc.). Data products include collections of in-situ data, reanalysis and trends of parameters, space and time aggregated in situ data and model outputs.

Biology

EMODnet Biology provides access to data on temporal and spatial distribution of marine species (angiosperms, benthos, birds, fish, macroalgae, mammals, phytoplankton, reptiles, zooplankton) and species traits from European regional seas as defined by the EEA's Europe's seas' dataset (Arctic Ocean, (North) Atlantic Ocean, Baltic Sea, Black Sea, Mediterranean Sea and North Sea). Measurements are focussed on biological occurrences: taxonomic identifications, the geotemporal information, the abundances/masses/counts. Other parameters provided include abiotic measurements (e.g. temperature, grain size), environmental facts (e.g. habitat), biotic measurements (e.g. abundance, length), biotic descriptors (e.g. lifestage, sex) and sampling descriptors (e.g. sampling instrument, surface area). Gridded products derived from the data in EMODnet are provided as well as the observation data itself.

2.2.1.3 Discovery and Access mechanism and functionality

There is a central portal available that unites all seven EMODnet thematic marine data services listed above into a single access point. The portal consists of three services:

- Map viewer: Further below the functionalities of the map viewer are described per theme.
 - <https://emodnet.ec.europa.eu/geoviewer/>
- Data Products catalogue: Further below the functionalities of the data product catalogue are described per theme.
 - <https://emodnet.ec.europa.eu/geonetwork/srv/eng/catalog.search#/home>
- ERDDAP:
 - <https://erddap.emodnet.eu/erddap/index.html>

Below for EMODnet Bathymetry, Chemistry, Physics and Biology more theme specific access options are also listed.

Bathymetry

For EMODnet Bathymetry there is a range of services and functionalities available to users for viewing and downloading bathymetry data products and for identifying and requesting access to the survey data sets that are used as basis input for the DTM. Currently the following key services and functionality are provided for users:

Theme-specific access option

- Data Discovery and Access service provides functionality to search and obtain survey data sets;
 - <https://cdi-bathymetry.seadatanet.org/search>

Central portal access option

- Data Products catalogue: Bathymetry Products Discovery and Access service provides functionality to search and view metadata of composite DTM;
- Map viewer & Data Products catalogue: High-Resolution DTM Discovery and Access: provides functionality to search and obtain HR-DTMs.
- Map viewer: service provides functionality to view, browse and download the DTM digital bathymetry and obtain information about the underlying data sets used to compile the DTM.

All Bathymetry layers are made available in the Map viewer and for many layers additional functionalities are offered such as depth profiles along user drawn transects, 3D viewing of the bathymetry in the browser, retrieving metadata of surveys, and downloading of DTM tiles and other data products.

Chemistry

A range of services and functionalities are provided to users for i) identifying and requesting access to the chemistry datasets that are gathered by EMODnet Chemistry, and ii) viewing, analysing, and downloading chemistry data products that are generated by EMODnet Chemistry:

Theme-specific access option

- Data Discovery and Access Service is the interface of SeaDataNet to search for, narrow down queries, and download selections of measurement data.
 - <https://cdi-chemistry.seadatanet.org/search>
- WebODV Data Explorer and Extractor allows users to create subsets, perform scientific analyses, generate various graphs, and download datasets in various formats from the validated aggregated data collections of EMODnet Chemistry.
 - <https://emodnet-chemistry.webodv.awi.de/>
- Marine litter tools are also available to view the distribution of beaches contained in the EMODnet Chemistry beach litter database, to generate EMODnet data formats for beach and seafloor litter, and to verify correctness.
 - <https://emodnet.ec.europa.eu/en/tools-guidelines>

Central portal access option

- Data Products catalogue: It provides functionality to search and browse metadata of chemistry data products and includes links to viewing and download.
- Map viewer: Allows users to view, browse, and interrogate EMODnet Chemistry data and data products. Specifically, EMODnet Chemistry provides the following map layers as OGC services:

Physics

Theme-specific access option

- ERDDAP EMODnet Physics
 - <https://erddap.emodnet-physics.eu/erddap/index.html>

Central portal access option

- Map viewer: It lets the user discover and access data per theme, platform, recording age, depth, provider and area. For each dataset/platform, a dedicated platform page is available. These pages provide the user with metadata, plots, download features, platform products, together with additional information and links, as well as statistics on the use of the data from that particular platform. Data quality information is available in connection with datasets, as well as the possibility to explore available machine-to-machine services.

Sextant

EMODnet data products are also available via the Sextant catalogue:

- <https://sextant.ifremer.fr/eng>

Here the user can start by searching on the metadata and then go on the view/download the data product. For more information you can see chapter: SeaDataNet Data Products & Sextant.

Biology

A range of services and functionalities are provided to users for i) identifying and requesting access to the datasets and ii) viewing, analysing, and downloading the generated data products:

Theme-specific access option

- The web services (metadata services, data visualisation services, data download services, non-OGC web services) are documented here:
 - <https://emodnet.ec.europa.eu/en/emodnet-web-service-documentation>

- The Integrated Publishing Toolkit ([IPT](#)), a freely available open source web application using the Darwin Core [standard](#), will make it easy to share biodiversity-related data and information with the EMODnet portal
- A [Quality Check tool](#) was created to support data providers in assessing the quality of their datasets prior to submission. It can be used either through an [Rshiny application or as an R package](#).

Central portal access option

- [Data Products catalogue](#): It provides functionality to search and browse metadata of EMODnet data products and includes links to viewing and download.
- [Map viewer](#): Allows users to view, browse, and interrogate EMODnet Biology data and data products.

EMODnet Biology implements (and if necessary adapts) common standards and vocabularies defined and used by SeaDataNet, WoRMS (World Register of Marine Species), OBIS (Ocean Biodiversity Information System), INSPIRE, GBIF (Global Biodiversity Information Facility), Marine Regions and the Lifewatch infrastructure.

2.2.2 Semantic Interoperability

2.2.2.1 Metadata format(s)

See the SeaDataNet CDI description in Chapter 4.4 for more information on the metadata formats.

2.2.2.2 Data format(s)

See the SeaDataNet CDI description in Chapter 4.4 for more information on the data formats.

2.2.2.3 Use of controlled vocabularies - which, where, how

See the SeaDataNet CDI description for more information on the controlled vocabularies. For EMODnet Chemistry use is made of the EMODnet Chemistry aggregated parameter names (P35) vocabulary, which is available at https://vocab.nerc.ac.uk/search_nvs/P35/.

2.2.3 Available web services for metadata access

The EMODnet web service documentation contains links to all EMODnet web services (CSW, WMS, WFS, WCS) that allow you to search, visualise and download EMODnet data and data products.

- <https://emodnet.ec.europa.eu/en/emodnet-web-service-documentation>

2.2.4 Available web services for data access

2.2.4.1 Data access policy - if yes, which and how deployed

Open access to data, no account required and free of charge.

2.2.4.2 Any web services and API's - URLs, function, how to operate

The EMODnet web service documentation contains links to all EMODnet web services (CSW, WMS, WFS, WCS) that allow you to search, visualise and download EMODnet data and data products.

- <https://emodnet.ec.europa.eu/en/emodnet-web-service-documentation>

2.3 Euro-Argo

2.3.1 Data Discovery and Access Component

The Euro-Argo portal features:

- The fleet monitoring tool that allows any user to visualise Argo profiling float metadata, ocean measurements, trajectories and technical parameters. It gives access to a fleet dashboard and also provides detailed information on a specific Argo float webpage (by WMO number).
- The Euro-Argo data selection tool that is specifically designed for users to select, visualise and download Argo scientific data (profiles files) in different formats.

There is a ERDDAP service available for retrieving Argo data as well. It is also possible to retrieve the whole Argo data collection by a DOI and associated landing page with descriptive metadata about the collection.

2.3.1.1 Name & Web Address

Euro-Argo: <https://www.euro-argo.eu/>

2.3.1.2 Types and number of data sets and/or data products

Argo collects salinity/temperature profiles as well as biogeochemical profiles that contain Oxygen, Nitrate, Chlorophyll-a, pH measurements, suspended particles and downwelling irradiance, from an array of robotic floats that populate the ice-free seas and oceans that are deeper than about 2000 m, even 4000 for few Argo floats.

Below, you can see an example of how an Argo float profile functions. A typical float cycle starts with the floater drifting at 1000 m depth, where every 10 days it then dives to about 2000 m depth, and starts measuring while the float is getting back up to the surface from that depth. When it reaches the surface, it sends the measurements to the corresponding Data Centres using satellite communications, before it then dives back to the drifting depth of 1000 m for a new cycle to start.

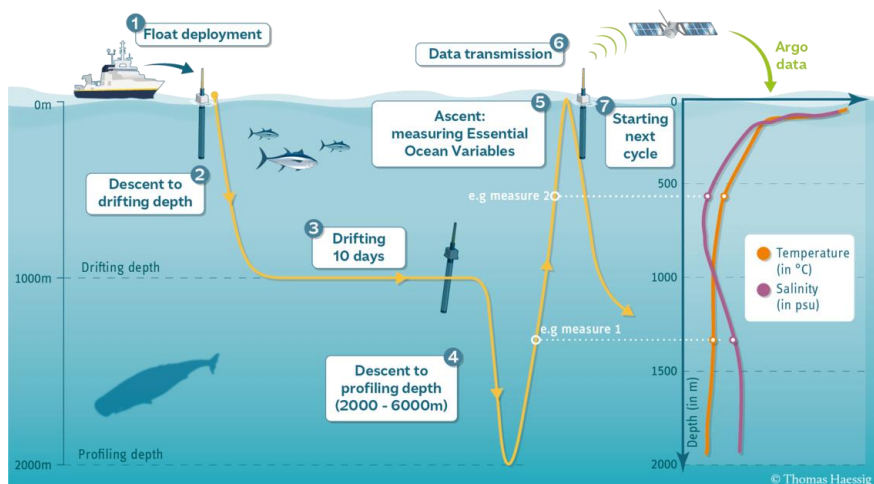


Figure 1 example of how an Argo float profile functions

Most profiles are made up of about 200 data points, but floats with high speed communications may be sending many more data points given the higher bandwidth. In total there are currently around 4000 Argo floats operational.

2.3.1.3 Discovery and Access mechanism and functionality

For discovery and access there are a number of mechanisms available to the user:

- Euro-Argo fleet monitoring tool: <https://fleetmonitoring.euro-argo.eu/dashboard>
 - The user gets access to a dashboard, where it is possible to make a selection of Argo floats in the left panel based on multiple filters, such as year of deployment, country, variable, etc. The resulting floats are then shown in a table in the middle panel with their respective latest floats positions shown on a global map in the right panel. It is possible to click on a float on the map and get more information including previous cycles with profile data, metadata, plots, and much more information. It is possible to download the profiles data in ASCII or NetCDF.
- Euro-Argo data selection tool: <https://dataselection.euro-argo.eu/>

2.3.2 Semantic Interoperability

2.3.2.1 Metadata format(s)

Each Argo float has a NetCDF (CF) metadata file with detailed information such as float type, serial number, scientist in charge of the float, scientist in charge of delayed mode adjustments, sensors, sensors serial numbers, mission, etc. The Argo metadata format is documented in the Argo user's [manual](#).

2.3.2.2 Data format(s)

Each Argo float has a NetCDF (CF) profile and trajectory file. The Argo profiles and trajectory formats are documented in the Argo user's [manual](#).

2.3.2.3 Use of controlled vocabularies - which, where, how

Argo physical parameters are linked to the [R03](#) vocabulary of the Nerc Vocabulary Service (NVS), with the corresponding units following the [P06](#) list. More information on the vocabularies used can be found in Argo user's [manual](#).

2.3.3 Available web services for metadata access

APIs used by the Euro-Argo web portals are open and publicly available to interested users at the following OpenAPI API endpoints (swagger):

- <https://fleetmonitoring.euro-argo.eu/swagger-ui.html>
- <https://dataselection.euro-argo.eu/swagger-ui.html>

2.3.4 Available web services for data access

2.3.4.1 Data access policy - if yes, which and how deployed

Open and free access, CC-BY 4.0 licensed.

2.3.4.2 Any web services and API's - URLs, function, how to operate

Beyond the mentioned OpenAPI API endpoints, there are as well available direct access mechanisms:

- Euro-Argo ERDDAP: <https://erddap.ifremer.fr/erddap/index.html>
- Euro-Argo ftp servers: <ftp://ftp.ifremer.fr/ifremer/argo/>
- Euro-Argo data collections:
 - <http://www.argodatamgt.org/Access-to-data/Argo-DOI-Digital-Object-Identifier>
 - <https://www.seanoe.org/data/00311/42182/>

2.4 SeaDataNet

2.4.1 SeaDataNet Common Data Index (CDI)

2.4.1.1 Data Discovery and Access Component

The SeaDataNet Common Data Index (CDI) service provides harmonised discovery and access to a large volume of marine and ocean data sets.

2.4.1.1.1 Name & Web Address

Common Data Index (CDI): <https://cdi.seadatanet.org/search>

2.4.1.1.2 Types and number of data sets and/or data products

The CDI service provides online access to a large number of marine and ocean data sets, managed by more than 117 connected SeaDataNet data centres originating from 34 countries around the European seas. Currently it gives access to more than 2.8 Million data sets, originating from more than 945 organisations in Europe, covering physical, geological, chemical, biological and geophysical data, and acquired in European waters and global oceans.

2.4.1.1.3 Discovery and Access mechanism and functionality

The online CDI User Interface includes functions for requesting access, and if granted, for downloading data sets from all connected data centres. The search function combines free search, facet search and geographic search options, powered by Elasticsearch, SQL search, and GeoServer.

Data requests can concern unrestricted and/or restricted data sets. Requests for unrestricted data sets are processed immediately after submission and requested data sets are made ready for download automatically from the SeaDataNet central unrestricted data cloud. While requests for restricted data sets are forwarded to the managers of connected data centres for their consideration, most of the cases are deliberated with data originators.

The processing of all data requests is controlled by the Request Status Manager (RSM) component which is integrated in the CDI User Interface. The RSM registers and processes all transactions, and communicates with the central unrestricted data cloud, users, and data centres. Users receive confirmation emails of their data set requests and subsequent processing, and can also check progress and undertake data downloading from the RSM service which is part of their personal dashboard.

2.4.1.2 Semantic Interoperability

2.4.1.2.1 Metadata format(s)

The CDI metadata gives information on the what, where, when, how, and who of each data set. It also gives standardised information on the data access restrictions that apply. The metadata format is based upon the ISO19115 content standard from TC211 and can be considered as a marine profile of the ISO19115 metadata content standard. The format supports SeaDataNet Cruise Summary Records (CSR), SeaDataNet Marine Data sets (EDMED), SeaDataNet Research Projects (EDMERP), and quality information.

The content and XML coding are following the ISO19139 schema and the CDI metadata format is compliant to the INSPIRE Directive Implementing Rules. SeaDataNet maintains and provides to data providers the [MIKADO](#) software tool to produce CDI XML files from local databases or local metadata files. Documentation about the CDI metadata format can be found [here](#).

2.4.1.2.2 Data format(s)

Delivery of SeaDataNet data sets to users is done by using, where possible, common SeaDataNet data transport formats, which interact with other SeaDataNet standards such as SeaDataNet controlled vocabularies, SeaDataNet European directories and SeaDataNet Quality Flag Scale as well as with SeaDataNet analysis and presentation tools such as ODV and DIVA. The following SeaDataNet data transport formats have been defined:

- SeaDataNet ODV4 ASCII for profiles, time series and trajectories
 - Can be used directly in the Ocean Data View (ODV) analysis and presentation software package, which is maintained and regularly extended with new functionalities.
- SeaDataNet NetCDF with CF compliance for profiles, time series and trajectories
 - This format has been defined by bringing together a community comprising NetCDF and CF experts (such as from NCAR and UNIDATA), and many users of oceanographic point data. This NetCDF format can be used as an alternative for the SeaDataNet ODV 4 ASCII format.
- SeaDataNet MedAtlas as optional extra format
- NetCDF with CF compliance for gridded data sets

SeaDataNet maintains and provides to data providers the [NEMO](#) software tool to convert from any type of ASCII format to the SeaDataNet ODV and MedAtlas ASCII formats as well as the SeaDataNet NetCDF (CF) format (for time series, profiles and trajectories observations) which are then made accessible through the CDI service. Another SeaDataNet OCTOPUS software tool is used by data providers as a multi-format checker, converter and splitter tool.

Next to these SeaDataNet common data formats, also a number of special SeaDataNet data formats are used and documented for specific data types, such as:

- SeaDataNet ODV ASCII format for biodiversity data as developed with EurOBIS
- SeaDataNet ODV ASCII format for micro litter data as developed with EMODnet Chemistry and TG-ML
- ASCII data format for beach litter data as developed with OSPAR, EMODnet Chemistry and TG-ML
- ASCII data format for seafloor litter data as developed with ICES, EMODnet Chemistry and TG-ML
- SeaDataNet ODV ASCII format for flowCytoMetry data;
- NetCDF4 (CF) format for HF Radar data as developed with EuroGOOS and EMODnet Physics

In addition, other common standards can be used for data formats. All applicable SeaDataNet data formats are included in the [L24](#) controlled vocabulary hosted by the British Oceanographic Data Centre (BODC) in the NERC Vocabulary Service (NVS). Documentation about the SeaDataNet data formats can be found [here](#).

2.4.1.2.3 Use of controlled vocabularies - which, where, how

The vocabulary services are technically managed and hosted by the British Oceanographic Data Centre (BODC) by means of the NERC Vocabulary Server (NVS). It includes the BODC parameter usage vocabulary ([P01](#)) and the SeaDataNet parameter discovery vocabulary ([P02](#)), which both include terms describing fine-grained groups of measurements used in the discovery interface. The BODC-approved storage units ([P06](#)) are used to label the units. Other lists include:

- [P24](#) - Units of measurement dimensions
- [P07](#) - Climate and Forecasting Standard Names
- [L05](#) - SeaDataNet Device Catalogue
- [L22](#) - SeaVoX Device Catalogue

2.4.1.3 Available web service for metadata access

The URLs for the CDI service endpoints are as follows:

- SPARQL service for CDI metadata
 - <https://cdi.seadatanet.org/sparql>
- OGC WMS service for locations of all data sets
 - <https://geo-service.maris.nl/seadatanet/wms?request=GetCapabilities>

- OGC WFS service for metadata features of all data sets
 - <https://geo-service.maris.nl/seadatanet/wfs?request=GetCapabilities>

The URLs for the SeaDataNet Common Vocabularies are as follows:

- SeaDataNet Common Vocabularies GUI
 - <https://vocab.nerc.ac.uk/>
- SeaDataNet Common Vocabularies REST API
 - <https://vocab.nerc.ac.uk/docs>
- SeaDataNet Common Vocabularies SPARQL
 - <https://vocab.nerc.ac.uk/sparql/>
- 1. SeaDataNet Common Vocabularies SOAP
 - 1.1. <http://vocab.nerc.ac.uk/vocab2.wsdl>

2.4.1.4 Available web service for data access

2.4.1.4.1 Data access policy - if yes, which and how deployed

All users can freely query and browse in the CDI directory; however, submitting requests for data access via the shopping basket requires that users are registered in the SeaDataNet central user register, thereby agreeing with the overall SeaDataNet User Licence.

2.4.1.4.2 Any web services and API's - URLs, function, how to operate

For the CDI service, SeaDataNet operates a web service which provides aggregated collections of CDIs at <https://cdi.seadatanet.org/report/aggregation/open>.

The related URLs are as follows:

- SeaDataNet OGC CSW endpoint:
 - <http://gs-service-production.geodab.eu/gs-service/services/essi/view/seadatanet/csw>
- with SeaDataNet CSW GetCapabilities:
 - <http://gs-service-production.geodab.eu/gs-service/services/essi/view/seadatanet/csw?service=CSW&request=GetCapabilities&version=2.0.2>
- SeaDataNet OAI-PMH endpoint:
 - <http://gs-service-production.geodab.eu/gs-service/services/essi/view/seadatanet/oaipmh>
- with SeaDataNet OAI-PMH Identify request:
 - <http://gs-service-production.geodab.eu/gs-service/services/essi/view/seadatanet/oaipmh?verb=Identify>
- OGC OpenSearch description endpoint:
 - <http://gs-service-production.geodab.eu/gs-service/services/essi/view/seadatanet/opensearch/description>
- With OpenSearch based demo portal:
 - <http://gs-service-production.geodab.eu/gs-service/search?view=seadatanet>

2.4.2 SeaDataNet Data Products & Sextant web catalogue

2.4.2.1 Data Discovery and Access Component

SeaDataNet provides aggregated datasets (ODV collections of all unrestricted SeaDataNet measurements of temperature and salinity by sea basins) and climatologies (regional gridded field products) based on the aggregated datasets. Each product is described in a Product Information Document (PIDoc) that can be accessed from the product's landing page.

The SeaDataNet page “Data Products” displays an embedded version of the Sextant web catalogue. The catalogue serves as a publishing and discovery service for data products. It allows producers and developers to publish metadata of their products and services while end-users can search for metadata according to different criteria (geographical area, theme, and producer).

2.4.2.1.1 Name & Web Address

SeaDataNet Data Products: <https://www.seadatanet.org/Products>

Sextant web catalogue: <https://sextant.ifremer.fr/eng>

2.4.2.1.2 Types and number of data sets and/or data products

The SeaDataNet Data Products include 25 data products of aggregated datasets and climatologies related to the marine environment.

The Sextant web catalogue contains 11,551 data products covering:

- Satellite, aerial and acoustic imaging
- Physics
- Biology
- Human activities
- Regulatory data
- Sea borders

2.4.2.1.3 Discovery and Access mechanism and functionality

SeaDataNet Data Products

The SeaDataNet Products catalogue includes an embedded interface to the Sextant metadata catalogue, showing all available SeaDataNet products. You can filter the products using text search or by drawing an area on the map to obtain products made with data located inside the drawn box or intersecting with it.

Each product is presented in an individual frame containing brief information about the product -title, overview, and source - and links to its online resources: download services, visualisation tools, and documentation.

You can access the products landing page via the DOI link present on both the frame of each product on the Data Products page and on each product description page. The DOI web page is a concise view of the product metadata and contains a link to the product PIDoc.

Sextant web catalogue

The Sextant web catalogue includes:

- Discovery User Interface for data products
- OGC Web Map Service for product visualisation

The products, published in Sextant, and displayed on the SeaDataNet portal, have an annexed unified documentation (Product Information Document, PIDoc) describing the methodology applied, informing about the product quality, and acknowledging the data sources and the tools used. The products obtain persistent Digital Object Identifier - DOIs, allowing their citation in scientific publications. The products are provided with web links to product data files, documentation and visualisation tools, i.e. from the product description in the catalogue, the user can jump to download and visualisation services.

2.5 GEBCO

2.5.1 Data Discovery and Access Component

The General Bathymetric Chart of the Oceans (GEBCO) aims to provide the most authoritative, publicly available bathymetry data sets for the world's oceans.

2.5.1.1 Name & Web Address

GEBCO web site, via one-click downloads for global coverage GEBCO grids:

- https://www.gebco.net/data_and_products/gridded_bathymetry_data/#global

Plus data download app for access to GEBCO grids for user-defined geographic areas:

- <https://download.gebco.net/>

2.5.1.2 Types and number of data sets and/or data products

Global grid of elevation values, in metres, on a 15 arc-second interval grid (a version is made available with ice elevation information for Greenland and Antarctica and one with sub-ice topography for these regions). Note that data from previous releases of the GEBCO grid are made available through the download app.

Type Identifier (TID) Grid – a global grid at 15 arc-second intervals providing information on type of data that the corresponding cell in the bathymetric/elevation grid is based on

- Grids of elevation values, in metres, for polar regions in polar stereographic projection coordinates (versions are made available with ice elevation information for Greenland and Antarctica and with sub-ice topography information for these regions).
- Imagery based on GEBCO grids

2.5.1.3 Discovery and Access mechanism and functionality

Data sets are lists as part GEBCO's data products on GEBCO's website:

- <https://www.gebco.net>
- https://www.gebco.net/data_and_products/

GEBCO's grids can be downloaded as a global file in netCDF or as tiled files in Esri ASCII raster and data GeoTiff formats. GEBCO's grids can be downloaded for user-defined areas through the download app. The user can enter the geographic coordinates directly or define a geographic area using the map. They can toggle to display the TID grid and to display the polar region grids and select data for polar regions in polar stereographic coordinates. Through the app, the user can select to download imagery based on the grids rather than downloading data.

2.5.2 Semantic Interoperability

2.5.2.1 Metadata format(s)

The netCDF file includes header information in CF-netCDF compliant format (1.6) and uses vocabs P01 and P06 to define parameters. The data set is accompanied by documentation in the format of a PDF file.

2.5.2.2 Data format(s)

- 2D netCDF
- GeoTIFF
- Esri ASCII

2.5.2.3 Use of controlled vocabularies - which, where, how

- P01 - BODC Parameter Usage Vocabulary, used to define parameters within the CF-netCDF file version of GEBCO's grids
- P06 - BODC-approved data storage units, used to define storage units for parameters defined within the CF-netCDF file version of GEBCO grids.

2.5.3 Available web services for metadata access

Not available to the best of our knowledge

2.5.4 Available web services for data access

2.5.4.1 Data access policy - if yes, which and how deployed

GEBCO licence (open access):

- https://www.gebco.net/data_and_products/gridded_bathymetry_data/gebco_2019/grid_terms_of_use.html

2.5.4.2 Any web services and API's - URLs, function, how to operate

The current GEBCO grid and TID grid (and previous versions of the grid) are available as Web Map Services:

- https://www.gebco.net/data_and_products/gebco_web_services/web_map_service/

2.6 NASA Earth Observation Data

2.6.1 Data Discovery and Access Component

The Earth Science Data Systems (ESDS) program governs the life cycle of Earth science data provided by NASA. The ESDS provides full access to NASA's collection of Earth science data.

2.6.1.1 Name & Web Address

Earth Data Open Access for Open Science - <https://www.earthdata.nasa.gov/>

2.6.1.2 Types and number of data sets and/or data products

The data includes remote sensing data, in-situ measurements and model outputs, focussing on the atmosphere, ocean, land, cryosphere and other topics. In the Earthdata search catalogue there are 9,333 data collections available to the user.

2.6.1.3 Discovery and Access mechanism and functionality

The Earthdata Search provides access to all of NASA's Earth science data holding via a user interface:

- <https://search.earthdata.nasa.gov/search>

It is possible to search for data collections using free text search as well as faceted search where filters can be applied based on the environmental topic, type of platform, organisations, projects, instruments and many more. Within the search catalogue, for each data collection there is relevant information provided as well as links to a unique landing page connected to that collection. It also includes links to services where the data collection can be downloaded, its unique DOI and documentation.

Once the user finds the relevant data collection, the data granules that are included within the collection are shown. Here it is now possible to search within the selected data collection on data granules by selecting a time period or using free text.

The data granules can be selected to show what region of the world they cover. It is possible for the data granules to get more descriptive information by either viewing the information or downloading the corresponding metadata. The data granules can be downloaded after the user has registered, where the provided data format depends on the data collection.

2.6.2 Semantic Interoperability

2.6.2.1 Metadata format(s)

Collection metadata is available in multiple formats:

- [ATOM](#)
- [CSV](#)
- [DIF-9](#)
- [DIF-10](#)
- [ECHO 10](#)
- [ISO-MENDS](#)
- [ISO-SMAP](#)
- [JSON](#)
- [UMM JSON](#)
- [KML](#)
- [Open Data](#)
- [STAC](#)
- [XML Reference](#)

The metadata accompanying the data granules can be downloaded in multiple [formats](#):

- Native
- UMM-G
- ATOM
- ECHO 10
- ISO19115

2.6.2.2 Data format(s)

There are many different data formats in which the data collections and respective data granules are provided. Depending on the collection, the available formats can be in NETcdf, GRIB, GeoTIFF, ASCII and many more.

2.6.2.3 Use of controlled vocabularies - which, where, how

Global Change Master Directory (GCMD) keywords, which contain a set of Earth Science vocabularies. See more information [here](#).

2.6.3 Available web service for metadata access

There is an Earthdata Developer Portal with information of relevant APIs that can be used to search and access Earth science data. More information on the relevant APIs can be found [here](#). Probably [CMR Search API](#) is a good entrypoint.

2.6.4 Available web service for data access

2.6.4.1 Data access policy - if yes, which and how deployed

Full and open data, users can get access to the data after registration.

2.6.4.2 Any web services and API's - URLs, function, how to operate

There is an Earthdata Developer Portal with information of relevant APIs that can be used to search and access Earth science data. More information on the relevant APIs can be found [here](#).

2.7 Copernicus Land Monitoring Service

Copernicus is the European Union's Earth observation programme. Information from this programme is provided through six thematic services: land, marine, atmosphere, climate change, emergency management and security. All information is free and openly accessible to all users.

2.7.1 Data Discovery and Access Component

2.7.1.1 Name & Web Address

The Copernicus Land Monitoring Service is available at: <https://land.copernicus.eu/>

2.7.1.2 Types and number of data sets and/or data products

The Land Service provides 246 products divided into four main components:

- Global: (65 products) provides a series of biogeophysical products on the status and evolution of the land surface at global scale at mid and low spatial resolution.
- Pan-European: provides information about land cover and land use and its changes, as well as biogeophysical parameters at European scale at high resolution
- Local: focuses on different hotspots, i.e. areas that are prone to specific environmental challenges and problems
- Imagery and reference data: satellite imagery forms the input for the creation of Land Monitoring products; and in order to ensure the efficient use of satellite imagery, in-situ data is required

Global products provide the following variables divided in four themes:

Theme	Variable
Vegetation	Land Cover, Fraction of photosynthetically active radiation absorbed by the vegetation, Fraction of green vegetation cover, Leaf Area index, Normalised Difference Vegetation Index, Vegetation Condition Index, Vegetation Productivity Index, Dry Matter Productivity, Burnt Area, Soil Water Index, Surface Soil Moisture
Water	Lake Water Quality, Water Bodies, Lake Surface Water Temperature
Energy	Land surface Temperature, Top of Canopy Reflectance, Surface Albedo
Cryosphere	Lake Ice Extent, Snow Cover Extent, Snow Water Equivalent

Pan-European products can be classified into:

- The Corine Land Cover inventory, consisting of an inventory of land cover in 44 classes.
- CLC+
- Pan-European High Resolution Layers (HRL), providing information on specific land cover characteristics
- Biophysical parameters
- European Ground Motion Service
- Related Pan-European products

The local component focuses on different hotspots, i.e. areas that are prone to specific environmental challenges and problems. The three local components are:

- Urban Atlas
- Riparian Zones
- Natura 2000 N2K

2.7.1.3 Discovery and Access mechanism and functionality

Users can browse the Copernicus Land Monitoring Service to find out the product of interest and learn more about them through the documentation. The portal is hierarchically structured, having four data macro categories corresponding to Global, Pan-European, Local, and Imagery and reference data.

Depending on the category the exact following steps to obtain metadata and data differ (different GUIs for global products and the others), but in general, once a product of interest is identified, it is possible to show it on a map, evaluate its metadata or download it at full (or a subset obtained by a second level search in case of global products). Credentials are required for download, they can be obtained by registering a personal account.

2.7.2 Semantic Interoperability

2.7.2.1 Metadata format(s)

The metadata format is reported to be an INSPIRE compliant profile of ISO 19115, its encoding is made according to XML (ISO 19139).

2.7.2.2 Data format(s)

Global products

NetCDF CF v. 1.6, GeoTIFF

Other products

ZIP package containing GeoTIFF files

2.7.2.3 Use of controlled vocabularies - which, where, how

Use in metadata

Controlled vocabulary	Vocabulary URL	Example terms	Comment
GEMET - INSPIRE themes, version 1.0	P22	Orthoimagery	thematic classification

Use in data

Controlled vocabulary	Vocabulary URL	Example terms	Comment
Climate and Forecast Standard Names	P07	surface_albedo	parameter

2.7.3 Available web services for metadata access

The following web services are available:

- Global catalogue:
 - <https://land.copernicus.vgt.vito.be/geonetwork/srv/eng/csw?>
- Pan-European, Local, Imagery and reference data catalogue:
 - <https://sdi.eea.europa.eu/catalogue/srv/eng/csw-copernicus?>

Both of the catalogues are OGC CSW v.2.0.2 services supporting the AP ISO profile, they are both implemented using the GeoNetwork technology.

2.7.4 Available web services for data access

A web service for complete machine to machine interaction seems not to be present.

Global products

The endpoint of an FTP service is sent to the user mail account after an order is performed by a human user on the global portal.

Other products

Maps can be visualised using OGC WMS, ArcGIS MapServer endpoints are available in the distribution information section of product metadata.

The download is human based, as the download link points to a web page where users can place an order.

2.7.4.1 Data access policy - if yes, which and how deployed

Credentials are required for download (registration of a personal user account). A quota of 500GB seems to be present for downloads from the global portal.

2.7.4.2 Any web services and API's - URLs, function, how to operate

Not available to the best of our knowledge

2.8 European Space Agency

2.8.1 Copernicus Open Access Hub

The Copernicus Open Access Hub (previously known as Sentinels Scientific Data Hub) provides complete, free and open access to Sentinel-1, Sentinel-2, Sentinel-3 and Sentinel-5P user products, starting from the In-Orbit Commissioning Review (IOCR).

2.8.1.1 Data Discovery and Access Component

2.8.1.1.1 Name & Web Address

Copernicus Open Access Hub is available at the web page: <https://scihub.copernicus.eu/>

2.8.1.1.2 Types and number of data sets and/or data products

Different user products are available from the Sentinel satellites for a total of 58,529,350 granules (scene data):

- **Level-0** products contain the compressed and unprocessed instrument source packets, with additional annotations and auxiliary information to support processing.
- **Level-1** focused data is the product intended for most data users and is generally available. The Level-0 product is transformed into a Level-1 product by the application of algorithms and calibration data to form a baseline engineering product, from which higher levels are derived.
- **Level-2** consists of geo-located geophysical products derived from Level-1.

The following list includes the available Sentinel products, grouped by satellite.

- [Sentinel-1](#):
 - **Level-0 Synthetic-Aperture Radar (SAR)**
 - The SAR Level-0 products consist of the sequence of Flexible Dynamic Block Adaptive Quantization (FDBAQ) compressed unfocused SAR raw data. For the data to be usable, it will need to be decompressed and processed using a SAR processor.
 - **Level-1 Single Look Complex (SLC), Level-1 Ground Range Detected (GRD)**
 - Level-1 Single Look Complex (SLC) products consist of focused SAR data geo-referenced using orbit and attitude data from the satellite and provided in zero-Doppler slant-range geometry. The products include a single look in each dimension using the full transmit signal bandwidth and consist of complex samples preserving the phase information.
 - Level-1 Ground Range Detected (GRD) products consist of focused SAR data that has been detected, multi-looked and projected to ground range using an Earth ellipsoid model. Phase information is lost. The resulting product has approximately square spatial resolution pixels and square pixel spacing with reduced speckle at the cost of worse spatial resolution.
 - **Level-2 Ocean (OCN)**
 - Ocean Wind field (OWI): The OWI is a ground range gridded estimate of the surface wind speed and direction at 10 m above the surface derived from internally generated Level-1 GRD images of SM, IW or EW modes.
 - Ocean Swell spectra (OSW): The OSW is a two-dimensional ocean surface swell spectrum and includes an estimate of the wind speed and direction per swell spectrum. The OSW is generated from Stripmap and Wave modes only. For Stripmap mode, there are multiple spectra derived from internally generated Level-1 SLC images. For Wave mode, there is one spectrum per vignette.

- Surface Radial Velocity (RVL): The RVL is a ground range gridded difference between the measured Level-2 Doppler grid and the Level-1 calculated geometrical Doppler.
- [Sentinel-2](#)
 - **Level 1-B Top-Of-Atmosphere radiances in sensor geometry:** One granule represents the sub-image one of the 12 detectors in the across track direction (25 km), and contains a given number of lines along track (approximately 23 km). Each Level-1B granule has a data volume of approximately 27MB.
 - **Level 1C Top-of-atmosphere reflectances in cartographic geometry:** Level-1C images are a set of tiles of 100 km² and approximately 500MB, and contains applied radiometric and geometric corrections (including orthorectification and spatial registration).
 - **Level 2-A Atmospherically corrected Surface Reflectances in cartographic geometry**
- [Sentinel-3](#)
 - **Ocean and Land Colour Instrument (OLCI):** OLCI products are generated in FR (Full Resolution 300 m) and in RR (Reduced Resolution 1.2 km) for the whole globe with the same coverage. A RR product is disseminated as a 'stripe' product (half orbit) and for FR as 'frames' (a few minutes along the orbit track). Level-1B products include Top-Of-Atmosphere (TOA) radiometric measurements in 21 spectral bands (400–1020 nm), radiometrically corrected, calibrated and spectrally characterised. It is ortho-geolocated (latitude and longitude coordinates, altitude) and annotated with satellite position and pointing, landmarks and preliminary pixel classification (e.g. land/water/cloud masks). The non-compressed size of a half orbit is 18.5GB for a FR (2.3GB for a RR). Level-2 products consist of geophysical quantities derived from the Level-1 product. Marine and land products are generated separately and contain extra atmospheric information relevant for both application domains. The non-compressed size of a half orbit is 28.4GB for a marine FR (2.4GB for a FR) and for land FR 8.8 GB (1.1GB for RR).
 - **Sea and Land Surface Temperature Radiometer (SLSTR):** Level-1B products include brightness temperatures and TOA Radiances ortho-geolocated and re-sampled onto the product grid. The measurement, annotation and auxiliary datasets are generated separately in two instrument views and at two resolutions: • 500 m resolution for solar reflectance bands (0.555, 0.659, 0.865, 1.375, 1.61 and 2.25 μm) • 1 km resolution for thermal infrared bands (3.74, 10.85 and 12 μm) • Level-2 products consist of marine and land geophysical quantities derived at 1 km from the Level-1 product.
 - **L1, L2 Synergy:** SYN Level- 2 data products are computed from OLCI Level-1B and SLSTR Level-1B data (including nadir and oblique views), and are composed of land and atmospheric geophysical quantities derived from these Level-1 products. VGT-P (1 km resolution) and SYN (300 m) products are disseminated as stripes (a full orbit). VGT-S are disseminated in tiles (global coverage or requested geographical subset).
 - **L0, L1, L2 Altimetry:** The SRAL instrument acquires data in two measurement modes:
 - Low-Resolution Mode (LRM): LRM mode is useful over open-ocean surfaces where topography is homogeneous.
 - Synthetic Aperture Radar (SAR) mode: to achieve high along-track resolution over relatively flat surfaces.

Sentinel-5P are currently separately available from the Sentinel-5P Pre-Operations Data Hub.

- Sentinel-5P

- **L1-B top of the atmosphere Earth radiances:** geo-located and radiometrically corrected top of the atmosphere Earth radiances in all spectral bands, as well as solar irradiances.
- **L2 products**
 - geolocated total columns of ozone, sulfur dioxide, nitrogen dioxide, carbon monoxide, formaldehyde and methane
 - geolocated tropospheric columns of ozone
 - geolocated vertical profiles of ozone
 - geolocated cloud and aerosol information (e.g. absorbing aerosol index and aerosol layer height)

2.8.1.1.3 Discovery and Access mechanism and functionality

Human users can search from the web portal GUI, formulating queries based on a set of constraints (spatial temporal extent, satellite platform, instrument, and other sensor related parameters).

The matching scenes (granules) overlays are displayed on the map, they can be subsequently downloaded. Sentinel-5P seems to be available at this time on a separate pre operational web portal.

2.8.1.2 Semantic Interoperability

2.8.1.2.1 Metadata format(s)

Atom feed responses are available at query time, giving back basic metadata elements: title, download url, summary, spatial temporal extent, ingestion date, fodmap, identifier, size, platform, instrument, and other Earth Observation attributes.

ISO 19139 metadata are only available upon data product download.

2.8.1.2.2 Data format(s)

Depending on the data products different data formats are available, usually NetCDF, JPEG 2000 or GeoTIFF files usually encapsulated in a SAFE container

2.8.1.2.3 Use of controlled vocabularies - which, where, how

To be investigated.

2.8.1.3 Available web service for metadata access

The Data Hub exposes two dedicated Application Program Interfaces (API) for browsing and accessing the EO data stored in the rolling archive:

- OData: OData is a data access protocol built on core protocols like HTTP and commonly accepted methodologies like REST , which can be handled by a large set of client tools (web browsers, download managers or computer programs such as cURL or Wget)
- OpenSearch: Open Search (Solr) is a set of technologies that allow publishing of search results in a standard and accessible format. It can be used to quickly identify the required resource which can then be downloaded by using OData. It is a RESTful technology and the Data Hub implementation uses the Apache Solr search engine

The following Open Search URI addresses the resource /Products exposed by the Open Search Service:

- https://scihub.copernicus.eu/dhus/search?q=*

Pagination, and sorting, are supported, along with search using the constraints also found in the web portal (e.g. platform name, ingestion date, spatial temporal, etc.).

Simple HTTP authentication is needed to use the OpenSearch endpoint (credentials can be obtained after personal account creation).

Correspondent queries to the Sentinel 5-P pre operational web portal (<https://s5phub.copernicus.eu/dhus/#/home>) fail with an unauthorised error message, hinting that separate credentials need to be requested to use this second portal.

2.8.1.4 Available web service for data access

2.8.1.4.1 Data access policy - if yes, which and how deployed

Credentials are needed for download (a personal account must be created). DHuS allows a maximum number of two simultaneous downloads attempts by a user.

2.8.1.4.2 Any web services and API's - URLs, function, how to operate

To download a full products OData service can be used, with the following syntax:

<ServiceRootUri>/Products('Id')/\$value

Example:

- [https://scihub.copernicus.eu/dhus/odata/v1/Products\('2b17b57d-fff4-4645-b539-91f305c27c69'\)/\\$value](https://scihub.copernicus.eu/dhus/odata/v1/Products('2b17b57d-fff4-4645-b539-91f305c27c69')/$value)

2.8.2 EUMETSAT

EUMETSAT is the European operational satellite agency for monitoring weather, climate and the environment from space. EUMETSAT operates the geostationary satellites Meteosat -10, and -11 over Europe and Africa, and Meteosat-9 over the Indian Ocean. It also operates two Metop polar-orbiting satellites as part of the Initial Joint Polar System (IJPS) shared with the US National Oceanic and Atmospheric Administration (NOAA). EUMETSAT is also a partner in the cooperative sea level monitoring Jason missions (Jason-3 and Jason-CS/Sentinel-6) involving Europe and the United States. The data and products from EUMETSAT's satellites are vital to weather forecasting and make a significant contribution to the monitoring of the environment and climate change.

The European Union has entrusted EUMETSAT with exploiting the four Sentinel missions of the Copernicus space component dedicated to the monitoring of atmosphere, ocean and climate on its behalf. EUMETSAT carries out these tasks in cooperation with ESA and already exploits the Sentinel-3 marine mission. EUMETSAT has established cooperation with operators of Earth observation satellites from Europe and China, India, Japan, South Korea and the United States. The cooperation with Russia was suspended in March 2022.

2.8.2.1 Data Discovery and Access Component

2.8.2.1.1 Name & Web Address

EUMETSAT homepage is at: <https://www.eumetsat.int/>

2.8.2.1.2 Types and number of data sets and/or data products

909 records are returned from 51 different platforms (satellites). The top ten platforms are:

- Metop (247)
- MSG (229)
- Sentinel-3 (74)
- MFG (65)
- NOAA (58)
- GOES (43)
- DMSP (40)
- Sentinel-6 (34)
- JPSS (26)
- Aqua (24)

2.8.2.1.3 Discovery and Access mechanism and functionality

Product navigator

The available products can be searched from the product navigator at: <https://navigator.eumetsat.int/start>. It acts as a showcase for the available products. Users can search by free text, temporal extent, and refine the query through the facet search (with the parameters: platform, sensor type, sensor, processing level, access, data policy, format, orbit, collection type, status, region, parameter, originating centre).

Once a product is identified, information on the access service that is serving the product is given to the user. Then, the user is redirected to a page listing all the different access mechanisms to proceed with the download (often requiring an order, after a second level search). Credentials are often required as well. These are:

EUMETSAT Data Store

Single point of access to search and download products from a growing catalogue of EUMETSAT's meteorological, climate and ocean data through Web User Interface (WebUI) and Application Programming Interfaces (API).

EUMETView

Online Map Service with functionalities for displaying visualisations and creating, animating and downloading maps from across the EUMETSAT product range.

EUMETSAT DATA TAILOR

Service for tailoring EUMETSAT data to user applications offering the capability to aggregate, filter layers, reproject, extract regions of interest, produce quick looks, and convert to popular and highly interoperable GIS formats.

EUMETSAT SFTP DATA ACCESS FOR COPERNICUS SENTINEL-6

Download service via the Internet for Copernicus Sentinel-6 data.

EUMETCAST SATELLITE

Access to near real-time data through DVB satellite with a guaranteed service level.

Available data: Meteosat, Metop, Jason, Copernicus Sentinel-3 data and third party products.

EUMETCAST TERRESTRIAL

Access to near real-time data through terrestrial networks. Restricted to agencies only.

Available data: Meteosat, Metop, Jason, Copernicus Sentinel-3 data and third party products.

DATA CENTRE

Ordering and delivery service for data with latency of more than 3 hours, historical and long-term archive data. Available data: Meteosat, Metop, Jason and Copernicus Sentinel-3 data.

2.8.2.2 Semantic Interoperability

2.8.2.2.1 Metadata format(s)

EUMETSAT metadata profile of ISO 19115-2 metadata records can be retrieved from the CSW

2.8.2.2.2 Data format(s)

Different data formats are available, the top ones being (as reported in the product navigator):

- netCDF (136), netCDF3 (22), netCDF4 (160)
- BUFR (192)
- HDF5 (146)

- GeoTIFF (105)
- SAFE (75)

2.8.2.2.3 Use of controlled vocabularies - which, where, how

2.8.2.3 Available web service for metadata access

Different web services are available to perform a search:

- OGC CSW endpoint, published at
 - <https://navigator.eumetsat.int/soapServices/CSWStartup> (909 records)
- REST API endpoint, published at
 - http://api.eumetsat.int/product-navigator/csw/record/_search/ (831 records)

APIs are as well present to search products (collections) available in the EUMETSAT data store, that is a subset (61 records) of all the products available through the EUMETSAT product navigator:

- The Data Store OpenSearch API
 - <https://eumetsatspace.atlassian.net/wiki/spaces/DSDS/pages/1660452893/The+Data+Store+OpenSearch+API>
- The Data Store Browse REST API
 - <https://eumetsatspace.atlassian.net/wiki/spaces/DSDS/pages/1660977169/The+Data+Store+Browse+REST+API>

Available Data Store collections are also listed here:

- https://api.eumetsat.int/data/browse/1.0.0/collections?format=html&access_token=e1a5d6c3-0382-33be-882d-ce94317b916c

2.8.2.4 Available web service for data access

Data is available through different publication mechanisms.

2.8.2.4.1 Data access policy - if yes, which and how deployed

Credentials are required in most cases to download the data.

2.8.2.4.2 Any web services and API's - URLs, function, how to operate

Products available through the EUMETSAT data store

APIs are available for download of those products that are available through the EUMETSAT data store. The Data Store Download REST API is available at:

- <https://eumetsatspace.atlassian.net/wiki/spaces/DSDS/pages/1660682293/The+Data+Store+Download+REST+API>

Products available through the EUMETview

WMS, WFS, WCS interfaces are as well available for those products:

- <https://view.eumetsat.int/geoserver/wms?service=WMS&version=1.3.0&request=GetCapabilities>
- <https://view.eumetsat.int/geoserver/ows?service=wfs&version=2.0.0&request=GetCapabilities>
- <https://view.eumetsat.int/geoserver/ows?service=WCS&version=2.0.1&request=GetCapabilities>

2.9 Climate Data Store

2.9.1 Data Discovery and Access Component

The Climate Data Store (CDS) is the cornerstone infrastructure that supports the implementation of the Copernicus Climate Change Service (C3S). It gives information about the past, present and future climate related to the ocean, atmosphere and land.

2.9.1.1 Name & Web Address

Climate Data Store: <https://cds.climate.copernicus.eu/cdsapp#!/home>

2.9.1.2 Types and number of data sets and/or data products

The CDS contains 18 applications and 142 datasets for the following product types:

- Climate indices
- Climate projection
- In-situ observations
- Reanalysis
- Satellite Observations
- Seasonal forecasts

It provides information regarding the [Essential Climate Variables](#) (ECVs), climate analyses, projections & reanalysis and indicators.

2.9.1.3 Discovery and Access mechanism and functionality

The CDS offers web-based search as well as API-based search and also provides a software toolbox that enables users to develop web-based applications using the datasets available in the CDS.

The Climate Data Store contains a GUI where the users can use text search to look for the relevant datasets with the possibility of faceted search by applying filters based on the product type, the domain (atmosphere, land or ocean), the spatial & temporal coverage and the sector.

Each data set has its own landing page, with general information of the dataset, description of the data, the variables that are included, links to documentation and a download page. Users can download data in three different ways, using the web interface, API or the toolbox.

Web interface

In the download page, there is a form where the user can select the variables, time period, region and format to download the file in. After submitting the form, the status of the request can be tracked in the 'your request' page. When it is 'complete', the result of the query can be downloaded.

The query that results from filling in the form can also be shown as an API request or as a toolbox request if for that specific dataset the toolbox is supported.

API

The CDS API allows for accessing the data in Python using the corresponding 'cdsapi' [package](#). The easiest way to get the relevant data in your Python environment is to use the web interface to construct the API query and then use that query in your Python environment with the package. More information can be found on this [page](#).

It must be noted that the API is REST-based meaning that it is not solely applicable for Python environments, but that it can also be extended to R for [example](#).

Toolbox

Here users can use the toolbox to make their own applications in an on-line interface using the data from the CDS. The toolbox is a catalogue of software with tools (statistics, averaging), workflows (output tool 1 → input tool 2) and applications (use workflows and datasets). For more information on the toolbox, check this [page](#).

2.9.2 Semantic Interoperability

2.9.2.1 Metadata format(s)

Each dataset has a web page with information on the data type, projection, coverage, resolution, file format, update frequency and variables.

2.9.2.2 Data format(s)

The data sets are in GRIB1, GRIB2 or NetCDF4 format. The GRIB formats use the WMO standards and the NetCDF format follows the CF-1.4, CF-1.5 or CF-1.6 convention. The Climate and Forecast Standard names conventions define metadata of what the data in each variable represents, and the spatial and temporal properties of the data.

2.9.2.3 Use of controlled vocabularies - which, where, how

Information is missing

2.9.3 Available web services for metadata access

Not available to the best of our knowledge

2.9.4 Available web services for data access

2.9.4.1 Data access policy - if yes, which and how deployed

The data is free and open, and in order to download the data, a registration is required.

2.9.4.2 Any web services and API's - URLs, function, how to operate

CDS Catalogue: <https://cds.climate.copernicus.eu/cdsapp#!/search>

2.10 NOAA

2.10.1 Physical Sciences Laboratory

2.10.1.1 Data Discovery and Access Component

The Physical Sciences Laboratory (PSL) contains a wide range of data, from gridded data sets spanning more than one hundred of years to NRT data continuously measured at one single location.

2.10.1.1.1 Name & Web Address

NOAA Physical Sciences Laboratory: <https://psl.noaa.gov/>

2.10.1.1.2 Types and number of data sets and/or data products

The PSL contains data sets related to the following themes:

- Climate
 - Gridded Climate Data
- Cloud Profiling Radars
- Air-Sea & Air-Land Fluxes
- Wind Profiling Radars
- Satellite
- Local Weather and Climate

2.10.1.1.3 Discovery and Access mechanism and functionality

Each of the above themes have sub-groups such as for example 'Gridded Climate Data' under the theme Climate. Each of these sub-groups have their own webpage with a specific way of accessing and downloading the related datasets.

In this analysis the focus will be on the Gridded Climate Data sets that can be found in:

- Gridded Climate Data Catalogue: <https://psl.noaa.gov/data/gridded/index.html>

The catalogue contains a GUI where the users can use text search to look for the relevant datasets with the possibility of faceted search by filtering on the category type (land, ocean, daily).

Each data set has its own webpage, with specifications of the dataset, such as the temporal/spatial coverage and more. It also includes a description of the data set, information about the usage of the data and references.

The dataset-specific web page includes a table with the variables stored in the dataset. By searching for the relevant variables in the data from the [table](#), there are multiple options to access the relevant data:

1. Access the data from the THREDDS catalogue by clicking on the book symbol next to the data you want to retrieve (see this [example](#)). In this example. the user can access the data [using](#):
 - a. OpenDAP, see this [page](#) for more information
 - b. HTTPServer, HTTP file download
 - c. WCS, which supports the access to geospatial data as coverages
 - d. WMS, which supports the access to georegistered map images from geoscience data sets
 - e. NETcdfSubset, which is a web service for subsetting CDM scientific grid datasets. In this [example](#), the user can subset the data on spatial/temporal coverage and levels.
2. Entering the sub-set [webpage](#) by clicking on the sub-set option, where the selected data can be filtered on region and date.
3. Clicking the download option.

Alternatively, the user can:

- Access the [THREDDS](#) catalogue directly and search for the relevant data there.
- Access the monthly composites plotting [tool](#) to download NCEP data.
- Use FTP, see this [page](#) for more information
- Use the PSL Climate Data [Repository](#)

2.10.1.2 Semantic Interoperability

2.10.1.2.1 Metadata format(s)

Each data set has a corresponding metadata file in XML using the ISO19115 metadata standard. There is also a NCML metadata representation available and an UDDC evaluation of how well the metadata contained in the data conforms to the NACDD (see for example this [page](#)).

2.10.1.2.2 Data format(s)

The data sets are in NetCDF4 format and follow the Climate and Forecast Standard names conventions that define the metadata of what the data in each variable represents, and the spatial and temporal properties of the data.

[2.10.1.2.3 Use of controlled vocabularies - which, where, how](#)
Information is missing

2.10.1.3 Available web service for metadata access

Not available to the best of our knowledge

2.10.1.4 Available web service for data access

2.10.1.4.1 Data access policy - if yes, which and how deployed

The data sets available via the PSL can be used freely by the public, unless specifically annotated otherwise.

[2.10.1.4.2 Any web services and API's - URLs, function, how to operate](#)
Not available to the best of our knowledge

2.10.2 National Centers for Environmental Information

2.10.2.1 Data Discovery and Access Component

The National Centers for Environmental Information (NCEI) manages a large archive of atmospheric, coastal, geophysical and oceanic research.

2.10.2.1.1 Name & Web Address

National Centers for Environmental Information: <https://www.ncei.noaa.gov/>

2.10.2.1.2 Types and number of data sets and/or data products

The NCEI archives more than 52 petabytes of data and offers users access to a wealth of Environmental Information (Biology, Physics, Weather, Coastal Indicators, etc.).

2.10.2.1.3 Discovery and Access mechanism and functionality

The NCEI includes multiple Discovery tools to enhance the Data Discovery and Access and interoperability.

In this analysis the focus will be on the NOAA OneStop tool, where data is discovered using the accompanying metadata record.

- NOAA OneStop tool: <https://data.noaa.gov/onestop/>

It contains 85,000 collection metadata records and over 24 million metadata granules. The user search interface contains a text search with some popular topics such as climate, oceans to guide the user. After searching for a specific keyword, the user can apply collection filters based on the location, date and attributes of the datasets. Once the correct dataset is found, a landing [page](#) appears with a general description of the dataset, link to a more descriptive webpage and multiple data access options such as ERDDAP, FTP, THREDDS, HTTPS and many more depending on the data set.

NCEI Geoportal:

- <https://www.ncei.noaa.gov/metadata/geoportal>
- 36,854 items can be found from the NCEI Geoportal

2.10.2.2 Semantic Interoperability

2.10.2.2.1 Metadata format(s)

The data set of the World Ocean Atlas 2018 has a corresponding metadata file in XML using the ISO19115-2 metadata standard. The metadata standards used in other data sets available on the NOAA OneStop tool might vary.

2.10.2.2.2 Data format(s)

There are multiple data formats available, namely ASCII, CSV, ArcGIS and NetCDF-4.

2.10.2.2.3 Use of controlled vocabularies - which, where, how

NODC vocabularies.

Climate and Forecasting Standard Names.

2.10.2.3 Available web service for metadata access

The CSW catalogue of the NCEI Geoportal is available at:

- <https://www.ncei.noaa.gov/metadata/geoportal/csw?>

2.10.2.4 Available web service for data access

2.10.2.4.1 Data access policy - if yes, which and how deployed

The data are openly available to the public.

2.10.2.4.2 Any web services and API's - URLs, function, how to operate

There is a Search API available that provides an endpoint for the NOAA OneStop data discovery. The documentation behind the API can be found [here](#) and [here](#).

2.11 European Soil Data Centre

The European Soil Data Centre (ESDAC) is the thematic centre for soil related data in Europe. Its ambition is to be the single reference point for and to host all relevant soil data and information at European level.

2.11.1 Data Discovery and Access Component

2.11.1.1 Name & Web Address

The European Soil Data Centre website is available at: <https://esdac.jrc.ec.europa.eu/>. The available datasets can be searched and browsed from the website at:

- <https://esdac.jrc.ec.europa.eu/resource-type/datasets>

2.11.1.2 Types and number of data sets and/or data products

The European Soil Data Centre (ESDAC) contains currently much soil data and information; most of the offered data are at European scale, while, when possible, links to national or global datasets are provided.

The ESDAC catalogue provides 98 datasets divided into the following five broad categories:

- [European Soil Database & soil properties](#)
- [Soil Threats Data](#) (erosion, soil organic carbon, landslides, compaction, salinization, soil biodiversity, contaminated sites, soil sealing, etc.)
- [Soil Functions Data](#)
- [Soil Point Data](#) (LUCAS, SPADE, etc)
- [Soil Projects Data](#) that contains data that stem from soil related [projects](#) at JRC

2.11.1.3 Discovery and Access mechanism and functionality

For discovery and access there is the following web page:

- <https://esdac.jrc.ec.europa.eu/resource-type/datasets>

Users can search and browse ESDAC datasets by:

- broad categories mentioned above
- themes: [Erosion by water](#), [Soil Organic Carbon Content](#), [Wind Erosion](#), [Gully Erosion](#), [Harvest Erosion](#), [Soil Compaction](#), [Soil Salinization](#), [Soil Biodiversity](#), [Landslides](#), [Soil Sealing](#), [Soil Contamination](#), [Soil Nutrients](#), [Soil pH](#), [Land degradation](#), [Desertification](#), [Soil Hydraulic Properties](#), [Soil Sampling](#), [Integrated Soil Modelling](#)
- networks and cooperations
- projects

There is also a handy one-page overview of all datasets publicly available in ESDAC at <https://esdac.jrc.ec.europa.eu/resource-type/datasets-list>, giving a short description of the datasets (name, link and short abstract).

Some datasets can be freely downloaded; others are accessible after prior registration, through a fill-in form.

There was also a Web map service available both as a web application (the ESDAC Map Viewer) and as a standardised Web Mapping Service (WMS), allowing discovery and access to European Soil Database and some additional data layers, e.g. related to main soil threats. But the service is out of order since June 2020:

- <https://esdac.jrc.ec.europa.eu/ApplicationAndServices>

2.11.2 Semantic Interoperability

2.11.2.1 Metadata format(s)

Each dataset is described by a detailed metadata page. Metadata fields are not fully harmonised between datasets. Here are the common elements: title, description, author temporal coverage, publications for citation, data access and use conditions. Here are the supplementary elements: Spatial Coverage, Resolution, Format, Projection.

No information if a metadata standard is used neither on the website nor in the related publications.

2.11.2.2 Data format(s)

Raster formats:

- GeoTIFF
- ESRI GRID

Vector formats:

- ESRI Shapefile

2.11.2.3 Use of controlled vocabularies - which, where, how

ESDAC has its own controlled vocabulary for themes relevant to soils (see the list of themes above) which is used as a criteria to search and browse the datasets available. However, this vocabulary is not published on the Web as FAIR terminology and it is not reused in the metadata of the dataset as keywords for example.

There is no controlled vocabulary used for describing the variables stored in the data.

2.11.3 Available web services for metadata access

There are no available web services for metadata access.

2.11.4 Available web services for data access

2.11.4.1 Data access policy - if yes, which and how deployed

All datasets are free to download. Some datasets are made accessible after prior registration, through a fill-in form asking for user details (name, e-mail, country, and affiliation). The request is manually processed.

Most ESDAC data can be used for any purpose with the only limitation not to distribute the data to third parties and with the requirement to cite the data properly when used in publications or data-products. These conditions of use are mentioned on each dataset metadata.

2.11.4.2 Any web services and API's - URLs, function, how to operate

The data download has to be done from each dataset webpage. For dataset requiring a prior registration through a fill-in form, the download url is sent by email. API development is considered for the future ([Panagos and al 2022](#)).

Ex: <https://esdac.jrc.ec.europa.eu/content/topsoil-soil-organic-carbon-lucas-eu25>

2.12 Joint Research Centre Data Catalogue

The Joint Research Centre (JRC) is the European Commission's in-house science service which employs scientists to carry out research in order to provide independent scientific advice and support to

policies of the European Union. The Data Catalogue contains an inventory of data produced by the JRC in accordance with the JRC Data Policy. The catalogue organises its content in dataset collections.

2.12.1 Data Discovery and Access Component

2.12.1.1 Name & Web Address

The Joint Research Centre Data Catalogue website is available at: <https://data.jrc.ec.europa.eu/>.

The available datasets can be searched from the web portal at: <https://data.jrc.ec.europa.eu/dataset>.

2.12.1.2 Types and number of data sets and/or data products

The JRC Data catalogue provides 3244 datasets divided into the the following science areas:

- Environment and climate change
- Safety and security
- Innovation and growth
- Nuclear safety and security
- Agriculture and food security
- Information society
- Energy and transport
- Standards
- Health and consumer protection
- Economic and monetary union

2.12.1.3 Discovery and Access mechanism and functionality

For discovery and access there is the following web portal:

- <https://data.jrc.ec.europa.eu/dataset>

Users can use free search to find the relevant data product and apply filters based on collections, contributors, keywords and science areas. It seems that most of the datasets are freely accessible and downloadable.

2.12.2 Semantic Interoperability

2.12.2.1 Metadata format(s)

The reference standard used for Metadata is DCAT-AP (DCAT application profile for data portals in Europe), that is actually the de facto EU metadata interchange format, extended in order to meet the identified requirements and needs for scientific data.

However, since JRC data are multi-disciplinary, and each discipline uses specific metadata standard, as well as specific controlled vocabularies for annotating metadata records, the JRC Data Catalogue infrastructure is designed to re-use as much as possible existing metadata, without requiring the original records to be re-created based on the reference standard (namely, [DCAT-AP](#)).

2.12.2.2 Data format(s)

To be investigated.

2.12.2.3 Use of controlled vocabularies - which, where, how

To be investigated.

2.12.3 Available web services for metadata access

One way to access the JRC Data Catalogue is to use the REST API. In particular, the following APIs are currently available:

- CKAN action API
- ODCAT action API

The documentation for these OPEN APIs is available at

- <https://data.jrc.ec.europa.eu/docs/index.html>.

Metadata can be retrieved in JSON or RDF formats.

CKAN action API

The **package_list** operation is available to query the catalogue and obtain the list of dataset identifiers:

- https://data.jrc.ec.europa.eu/api/3/action/package_list

The **package_show** operation is used to query the catalogue and obtain the metadata (JSON format) for a specific dataset. For example:

- https://data.jrc.ec.europa.eu/api/3/action/package_show?id=0cf7f0c0-b04a-4e57-a45b-e13be8bdf5e4

ODCAT action API

The **collection_list** operation is available to obtain the dataset collection list:

- https://data.jrc.ec.europa.eu/api/1/action/collection_list

The **collection_show** and **dataset_show** are used to query the catalogue and get the metadata for a specific collection:

- https://data.jrc.ec.europa.eu/api/1/action/collection_show?collection_id=airmex

or a specific dataset:

- https://data.jrc.ec.europa.eu/api/1/action/dataset_show?dataset_id=jrc-airmex-campaigns-data

2.12.4 Available web services for data access

2.12.4.1 Data access policy - if yes, which and how deployed

The JRC prepared a dedicated Data Policy to complement the JRC Policy on Open Access to Scientific Publications and Supporting Guidance to promote open access to research data in the context of Horizon 2020.

2.12.4.2 Any web services and API's - URLs, function, how to operate

Direct Download

In many cases direct data download links can be obtained following the URLs reported in the transfer options. This holds for example whenever URLs are given according to the template:

https://data.jrc.ec.europa.eu/dataset/{dataset_identifier}

For example:

- <https://data.jrc.ec.europa.eu/dataset/e83b19ce-08c2-4e0c-b93a-5fd62be21e5e>

Downloading the link an html page is obtained, that contains a set of direct link URLs.

2.13 European Environment Agency

The European Environment Agency (EEA) is an agency of the European Union, whose task is to provide sound, independent information on the environment. The EEA aims to support sustainable development by helping to achieve significant and measurable improvement in Europe's environment, through the provision of timely, targeted, relevant and reliable information to policy making agents and the public.

2.13.1 Data Discovery and Access Component

2.13.1.1 Name & Web Address

European Environment agency website is available at: <https://www.eea.europa.eu/>

2.13.1.2 Types and number of data sets and/or data products

A search for data + interactive data in the EEA website gives back 1283 results.

A search from EEA SDI gives back 2229 datasets. Note: Copernicus Land Monitoring Service Pan-European products are also provided here.

2.13.1.3 Discovery and Access mechanism and functionality

Users can browse EEA site structured by themes (air and climate, nature, sustainability and well-being, economic sectors) to find out datasets of interest. Otherwise a search can be performed, leveraging two different portals:

- EEA website: <https://www.eea.europa.eu/data-and-maps>
- EEA SDI: <https://sdi.eea.europa.eu/>

The first portal contains resources of different types (e.g. Article, Data, Graph, Indicator, Infographic, Interactive data, Interactive map, Map, News, Publication, Video), while the second one is focused on spatial data.

Focusing on the SDI, it is possible to browse datasets by two different categorizations:

- INSPIRE themes
- EEA topics

It's also possible to search on a map or by keyword.

2.13.2 Semantic Interoperability

2.13.2.1 Metadata format(s)

The EEA metadata profile (described at this [web page](#)) is used to describe datasets. The EEA metadata profile follows the Technical Guidance for the implementation of INSPIRE data set and service metadata based on ISO/TS 19139:2007 which is the implementation specification for defining metadata for INSPIRE data sets and services in ISO/TS 19139 based XML format in compliance with the INSPIRE IR for metadata and network services (Regulation 1205/2008 and Regulation 1089/2010). The EEA metadata profile considers as compulsory all optional elements offered by ISO 19115 and INSPIRE which are regarded as useful for EEA activities.

Example metadata record:

- <https://sdi.eea.europa.eu/catalogue/srv/api/records/a0aacdd2-46b2-485a-ac49-162936d84395/formatters/xml>

4.13.2.2 Data format(s)

Multiple formats are expected from EEA according to EEA metadata profile:

Raster formats:

AAIGrid	Arc/Info ASCII Grid
AIG	Arc/Info Binary Grid (.adf)
BIL	Binary interleaved (.BIL)
CEOS	CEOS format (used by some satellite image providers)
DIMAP	DIMAP format (used by some satellite image providers)
ECW	ERMapper Compressed Wavelets (.ecw)
Ehdr	ESRI .hdr Labelled
ENVI	ENVI .hdr Labelled Raster
GRASS	GRASS Rasters
GTiff	TIFF / BigTIFF / GeoTIFF (if no georeferencing information is available in the TIFF file itself, an ESRI world file [.tfw] must be present)
HDF	Hierarchical Data Format Release (specify the version number [4 or 5] in the metadata element version)
HFA	Erdas Imagine (.img)
LAN	Erdas 7.x .LAN and .GIS
MrSID	Multi-resolution Seamless Image Database
netCDF	NetCDF
XYZ	ASCII Gridded XYZ

Vector formats:

AVCbin	Arc/Info Binary Coverage
AVCE00	Arc/Info .E00 (ASCII) Coverage
CSV	Comma Separated Value (.csv)
FGeo	ESRI File GeoDatabase
GML	Geography Markup Language
GRASS	GRASS vector
KML	Keyhole Markup Language
MIF	!MapInfo File

PGeo	ESRI Personal GeoDatabase
SHP	ESRI Shapefile
Postgis	Postgresql/Postgis database
Spatialite	SQLite/Spatialite database

The following is an occurrence analysis in the first 100 metadata records. It highlights the heterogeneity of formats:

- 17 | GeoTIFF
- 14 | unknown
- 13 | ascii (.csv, .txt, .sql)
- 11 | SHP
- 8 | Microsoft Excel (.xls, .xlsx)
- 7 | Microsoft Access (.mdb, .accdb)
- 6 | Geopackage
- 4 | AAIGrid
- 4 | PGeo
- 3 | ESRI Shapefile
- 3 | FGeo
- 3 | WMS Dienst
- 2 | Geopackage (ogc)
- 2 | WFS Dienst
- 1 | AIG
- 1 | DXF
- 1 | DXF/DWG
- 1 | ESRI Personal Geodatabase
- 1 | ESRI shapefile
- 1 | INTERLIS
- 1 | Interlis 2
- 1 | Shapefile
- 1 | SHP-ArcView ShapeFile

In general the downloaded file will be a ZIP package containing the desired file(s).

2.13.2.3 Use of controlled vocabularies - which, where, how

Use in metadata

Controlled vocabulary	Vocabulary URL	Example terms	Comment
GEMET - INSPIRE themes, version 1.0	P22	Protected sites	thematic classification
EEA topics	http://www.eea.europa.eu/portal_vocabularies/themes	Biodiversity - Ecosystem	thematic classification
INSPIRE Spatial scope	https://inspire.ec.europa.eu/metadata-codelist/SpatialScope	European	spatial scope

Other vocabularies are as well used according to the EEA metadata profile: “EEA keywords list”, “EEA data categories”, “Continents, countries, sea regions of the world”, although their definitions couldn’t be easily found.

Use in data

Needs to be investigated

2.13.3 Available web services for metadata access

EEA SDI catalogue is an INSPIRE compliant Discovery Service which follows the OGC CSW 2.0.2 standard. It is implemented using the open source catalogue software GeoNetwork.

Its endpoint is <https://sdi.eea.europa.eu/catalogue/srv/eng/csw>

2.13.4 Available web services for data access

2.13.4.1 Data access policy - if yes, which and how deployed

In some cases credentials are required to proceed to download

2.13.4.2 Any web services and API's - URLs, function, how to operate

Direct download

In many cases direct data download links can be obtained following the URLs reported in the transfer options. This holds for example whenever URLs are given according to the template:

- <https://sdi.eea.europa.eu/data/{dataset identifier}>

For example: <https://sdi.eea.europa.eu/data/61dcacc3-76f3-4ed0-af9a-c628fa4d206a>

Downloading the link an html page is obtained, that contains a set of direct link URLs.

WMS services

WMS endpoints are also often referenced in the metadata records. For example:

- <http://climate-adapt.eea.europa.eu/geoserver/wms>

2.14 National Center for Biotechnology Information (NCBI)

The National Center for Biotechnology Information (NCBI) Taxonomy includes organism names and classifications for every sequence in the nucleotide and protein sequence databases of the International Nucleotide Sequence Database Collaboration (INSDC). It provides a framework for clustering elements within other domains of NCBI web pages, for internal linking between domains of the Entrez system and for linking out to taxon-specific external resources on the web and relevant publications. It is also the standard nomenclature and classification repository for the International Nucleotide Sequence Database Collaboration (INSDC) that comprises GenBank, the European Molecular Biology Laboratory (EMBL) and DNA Data Bank of Japan (DDBJ). It consists of a curated set of names and classifications for all of the source organisms represented in the INSDC.

2.14.1 Data Discovery and Access Component

2.14.1.1 Name & Web Address

This is the link to the taxonomy browser:

- <https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi>

This has been converted to an ontology by the [OBO Foundry](#) here:

- <https://ontobee.org/>

2.14.1.2 Types and number of data sets and/or data products

Taxon names

The NCBI Taxonomy database (<http://www.ncbi.nlm.nih.gov/taxonomy>) is the standard nomenclature and classification repository for the International Nucleotide Sequence Database Collaboration (INSDC), comprising the GenBank, ENA (EMBL) and DDBJ databases.

There is a link between the taxon IDs in WoRMS (AphiaIDs) and the NCBI IDs: where it is known that there is a match between the two, both databases contain each other's IDs.

2.14.1.3 Discovery and Access mechanism and functionality

Under the NCBI Taxonomy Browser two different kinds of web pages are supported. Hierarchy pages present the taxonomic classification, while taxon-specific pages summarise all of the information associated with a particular taxonomic entry. By default, the hierarchy displays three levels in the classification, but this can be changed (zero levels display the taxon-specific page). The hierarchy pages are also customised to display hotlinked counts of entries in other Entrez databases. Taxon specific pages will display the names and their relational terms associated with that entry (except for misspellings and unpublished names). The lineage can display a full or abbreviated classification.

In addition to this, manually curated information is displayed as well. This includes type material and comments annotated by the taxonomic curators as well as relevant literature and type material information, with hotlinks as appropriate. An Entrez records table shows links to other Entrez databases, including links to records from type material, when they are available and annotated. At the bottom of the taxon specific pages additional layered information sets are found. For relevant organisms this can include Genome information with links to resource information in other databases.

Another set of links are External Information Resources or NCBI LinkOut. This provides outside groups the ability to maintain links to externally curated resources after submitting information for their web-accessible resources (<https://www.ncbi.nlm.nih.gov/projects/linkout/>). Another block below this will display the modifiers such as strain, isolate or culture collection associated with the organism in GenBank sequence entries. Finally, additional search capabilities are supported in the Taxonomy Browser such as a wild card search.

2.14.2 Semantic Interoperability

2.14.2.1 Metadata format(s)

Ontology Web Language (owl) and RDF

2.14.2.2 Data format(s)

Only for the OBO foundry version of the NCBI, it is an ontology that uses owl and RDF. The taxonomy browser is in csv, ascii and html.

2.14.2.3 Use of controlled vocabularies - which, where, how

Only for the OBO foundry version of the NCBI, it is an ontology that uses owl and RDF. The taxonomy browser is in csv, ascii and html.

2.14.3 Available web services for metadata access

The NCBI taxon DB has been implemented as an ontology by OBO Foundry. To get machine readable information the OBO Foundry ontology has to be used. Each taxonid is identified using this URL:

- https://ontobee.org/ontology/rdf/NCBITaxon?iri=http://purl.obolibrary.org/obo/NCBITaxon_ID
- e.g. https://ontobee.org/ontology/rdf/NCBITaxon?iri=http://purl.obolibrary.org/obo/NCBITaxon_11103

2.14.4 Available web services for data access

NCBI browser generally offers ascii, csv or html. For machine actionable info use the above. Another service was found outside of NCBI that offers a json output using REST, for example:

- <https://rest.ensembl.org/taxonomy/id/1386?content-type=application/json>

2.14.4.1 Data access policy - if yes, which and how deployed

The NCBI taxonomy database is not a primary source for taxonomic or phylogenetic information. Furthermore, the database does not follow a single taxonomic treatise but rather attempts to incorporate phylogenetic and taxonomic knowledge from a variety of sources, including the published literature, web-based databases, and the advice of sequence submitters and outside taxonomy experts. Consequently, the NCBI taxonomy database is not a phylogenetic or taxonomic authority and should not be cited as such.

However, if you want to acknowledge the fact that you have used the NCBI taxonomy database and other NCBI services and databases for your research, the two following papers can be referenced:

- Schoch CL, et al. NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database (Oxford). 2020: baaa062. [PubMed]
- Sayers EW, et al. GenBank. Nucleic Acids Res. 2019. 47(D1):D94-D99. [PubMed]

2.14.4.2 Any web services and API's - URLs, function, how to operate

UI: <https://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/index.cgi>

Ontology:

- https://ontobee.org/ontology/rdf/NCBITaxon?iri=http://purl.obolibrary.org/obo/NCBITaxon_ID
- E.g.
 - https://ontobee.org/ontology/rdf/NCBITaxon?iri=http://purl.obolibrary.org/obo/NCBITaxon_11103

2.15 Global Biodiversity Information Facility (GBIF)

GBIF—the Global Biodiversity Information Facility—is an international network and data infrastructure funded by the world's governments and aimed at providing anyone, anywhere, open access to data about all types of life on Earth.

2.15.1 Data Discovery and Access Component

2.15.1.1 Name & Web Address

The Global Biodiversity Information Facility (GBIF) homepage is online at <https://www.gbif.org/>

2.15.1.2 Types and number of data sets and/or data products

83,079 datasets (collections) comprising hundreds of millions (2,313,730,034) of species occurrence records.

2.15.1.3 Discovery and Access mechanism and functionality

Users can search through the web portal occurrences, species, datasets, publishers and general resources. Depending on the resource type, different search constraints are available. Once an interesting resource is identified, depending on the resource type, it is possible to proceed to the download in different data formats.

2.15.2 Semantic Interoperability

2.15.2.1 Metadata format(s)

Ecological Metadata Language (EML)

2.15.2.2 Data format(s)

Darwin Core Archive

2.15.2.3 Use of controlled vocabularies - which, where, how

The DwC-A terms are used in the DwC-A files, and all “measurements and facts” are linked to controlled vocabularies: which ones are used depends on the data and the data creator; those from the ENVO ontology, and BODC, WoRMS, and Marine Regions for the ocean-linked data, are certainly used, as are MiXS and NCBI (and other omics) IDs for DNA-based data.

2.15.3 Available web services for metadata access

GBIF Registry API can be used to obtain information about datasets (collections).

For example the following operations are available:

- Lists all datasets (<https://api.gbif.org/v1/dataset>)
- Gets details for the single dataset
 - <https://api.gbif.org/v1/dataset/4fa7b334-ce0d-4e88-aaae-2e0c138d049e>

2.15.4 Available web services for data access

2.15.4.1 Data access policy - if yes, which and how deployed

The following licences are found:

- CC0 1.0 (1,673 datasets)
- CC BY 4.0 (27,956 datasets)
- CC BY-NC 4.0 (4,873 datasets)

2.15.4.2 Any web services and API's - URLs, function, how to operate

Full dataset

A direct link to download the full Darwin Core Archive of an entire dataset is available in the response of GBIF Registry API. Example given:

- <https://hosted-datasets.gbif.org/eBird/2021-eBird-dwca-1.0.zip>

Occurrences

Occurrences can be obtained and downloaded from the GBIF Occurrence API

2.16 Ocean Biodiversity Information System (OBIS)

OBIS is a global open-access data and information clearing-house on marine biodiversity for science, conservation and sustainable development. More than 20 OBIS nodes around the world connect 500 institutions from 56 countries. EurOBIS is the European node: the data published on EurOBIS are harvested by OBIS; and the services and products served by EMODnet Biology are built on EurOBIS data. OBIS emanates from the Census of Marine Life (2000-2010) and was adopted as a project under IOC-UNESCO's International Oceanographic Data and Information (IODE) programme in 2009. OBI joined GBIF as an associate participant in 2001 and became the primary marine data provider to GBIF in 2004.

2.16.1 Data Discovery and Access Component

2.16.1.1 Name & Web Address

Ocean Biodiversity Information System (OBIS) homepage is available at <https://obis.org/>

2.16.1.2 Types and number of data sets and/or data products

4601 datasets (collections) including 45 million observations of nearly 120 000 marine species, from Bacteria to Whales, from the surface to 10,900 metres depth, and from the Tropics to the Poles.

2.16.1.3 Discovery and Access mechanism and functionality

Search by species name, higher taxonomic level, geographic area, depth, time and environmental parameters is supported.

Resulting occurrences are visualised on a map, along with a statistical analysis. The top datasets contributing to the occurrences result are also referenced from the result page. By clicking on each dataset more information on the specific dataset can be obtained, including links to download the dataset.

2.16.2 Semantic Interoperability

2.16.2.1 Metadata format(s)

EML, RTF

Example EML: https://www.marine.csiro.au/ipt/eml.do?r=imos_atf_ad&v=1.9

Example RTF: https://www.marine.csiro.au/ipt/rtf.do?r=imos_atf_ad&v=1.9

2.16.2.2 Data format(s)

Individual datasets

Darwin Core Archive (DwC-A)

Example DwC-A: https://www.marine.csiro.au/ipt/archive.do?r=imos_atf_ad&v=1.9

Full OBIS archive

Full export of OBIS archive is available as csv and parquet data formats.

2.16.2.3 Use of controlled vocabularies - which, where, how

Information is missing

2.16.3 Available web services for metadata access

The OBIS REST API (based on a JSON data model) is available at: <https://api.obis.org/>

Different resources are available, the main ones being:

- Occurrences: find and fetch occurrences by a set of query constraints

- Taxon: get taxon record hierarchy
- Dataset: find dataset (collections) records by a set of query constraints
- Institute: find institute records
- Area: find area records
- Country: find country records
- Statistics: Get statistics about occurrence records

2.16.4 Available web services for data access

2.16.4.1 Data access policy - if yes, which and how deployed

Probably dataset dependant, in many cases is CC-BY-NC

2.16.4.2 Any web services and API's - URLs, function, how to operate

Direct link to entire datasets

The results obtained from the Dataset operation of the OBIS REST API contains a link to direct download the dataset. The resulting file is a zip file containing EML and RTF metadata as well as DwC-A data.

Granules download

Occurrences data from one or more datasets can be obtained using the Occurrences operation in the OBIS REST API.

2.17 European Nucleotide Archive (ENA)

2.17.1 Data Discovery and Access Component

The European Nucleotide Archive (ENA) provides a comprehensive open record of the world's nucleotide sequencing information and a platform for the management and analysis of sequence and related data. Covering raw sequencing data, sequence assembly information, functional annotation and a host of further data types, content is measured in millions of taxa, hundreds of thousands of sequenced libraries and petabytes of storage. ENA is operated by the EMBL European Bioinformatics Institute (EMBL – EBI). ENA is designated by the ELIXIR infrastructure both as a Core Data Resource, and a Deposition Database.

2.17.1.1 Name & Web Address

ENA Browser: <https://www.ebi.ac.uk/ena/browser/home>

2.17.1.2 Types and number of data sets and/or data products

ENA covers many data types in a number of interlinked database tables. It contains 2×10^9 sequences and 1×10^{16} base pairs of read data across 2×10^6 taxa.

2.17.1.3 Discovery and Access mechanism and functionality

The ENA browser brings together a set of services via web interfaces, built upon underlying APIs. Two relevant services are:

- data discovery (metadata search and retrieval)
 - <https://www.ebi.ac.uk/ena/browser/advanced-search>
- data retrieval
 - <https://www.ebi.ac.uk/ena/browser/home>

Search queries can be built using UI wizard at <https://www.ebi.ac.uk/ena/browser/advancedsearch> and run in the Browser, or using the “Copy Curl Request” button exported as a curl command to be run in the command line. There are many criteria to build a query. A good guidance can be found at:

- <https://ena-browser-docs.readthedocs.io/en/latest/browser/search/advanced.html#search-query>

2.17.2 Semantic Interoperability

2.17.2.1 Metadata format(s)

Indexed metadata is available as TSV (Tab Separated Value) by default or as JSON (format=json). Details can be found at:

- <https://ena-browser-docs.readthedocs.io/en/latest/browser/search/advanced.html#downloadresults-report>

2.17.2.2 Data format(s)

Data can be retrieved in different formats and with easy file download options through RESTful services:

- EMBL Flatfile format
- FASTA format for sequences
- XML Format

Details about formats:

- <https://ena-browser-docs.readthedocs.io/en/latest/browser/search/advanced.html#downloadena-records>

2.17.2.3 Use of controlled vocabularies - which, where, how

Controlled Vocabularies are used and support building a search query as can be seen in the Query Builder component of:

- <https://www.ebi.ac.uk/ena/browser/advanced-search>

Also, it can be seen using the controlled Vocabularies endpoint by passing in a field name:

- <https://www.ebi.ac.uk/ena/portal/api/controlledVocab?field=checklist>

2.17.3 Available web services for metadata access

Relevant API's are:

ENA Data Discovery: <https://www.ebi.ac.uk/ena/portal/api/>

How to use the API's and build machine-to-machine services can be found in the documentation of the ENA Portal API: <https://www.ebi.ac.uk/ena/portal/api/doc>

2.17.4 Available web services for data access

2.17.4.1 Data access policy - if yes, which and how deployed

Terms of use for web and RESTful services can be found at:

- <https://www.ebi.ac.uk/about/terms-of-use/>

GDPR notice for ENA Browser:

- <https://www.ebi.ac.uk/data-protection/privacy-notice/ena-presentation>

GDPR notice for both RESTful services:

- <https://www.ebi.ac.uk/data-protection/privacy-notice/embl-ebi-public-website>

2.17.4.2 Any web services and API's - URLs, function, how to operate

Relevant API's are:

- ENA Data Retrieval: <https://www.ebi.ac.uk/ena/browser/api/>

How to use the API's and build machine-to-machine services can be found in the documentation of the ENA Portal API:

- <https://www.ebi.ac.uk/ena/portal/api/doc>

Data downloading (retrieving sequences) using command-line, tools, FTP:

- <https://ena-docs.readthedocs.io/en/latest/retrieval/file-download.html>

2.18 World Register of Marine Species

2.18.1 Data Discovery and Access Component

The World Register of Marine Species provides an authoritative and comprehensive list of names of marine organisms, including information on synonymy, validity status, taxonomic classification. It can serve as a guide to interpret taxonomic literature, and uniquely reference and validate taxonomic names.

2.18.1.1 Name & Web Address

World Register of Marine Species (WoRMS) website: <https://www.marinespecies.org/>

2.18.1.2 Types and number of data sets and/or data products

As of March-2023, WoRMS contains 785,350 taxonomic names including some for fossil organisms. 97% of the names have been checked by taxonomic experts. Of the 643,591 species names, 338,206 are accepted species names. Besides species names, Aphia also contains the higher classification in which each scientific name is linked to its parent taxon. All names have PIDs (AphiaIDs) which guarantee machine-interoperability and accessibility, as well providing semantic annotation for taxonomic data. Other information is provided for the names in WoRMS: distributions, literature citations, indicators, traits, etc.

2.18.1.3 Discovery and Access mechanism and functionality

WoRMS provides web-based services for searching in its database of taxonomic and associated information, and for running taxonomic checks on names uploaded in CSV files. Web Services (REST, SOAP; <https://www.marinespecies.org/aphia.php?p=webservice>) are also provided to feed your own applications with standard WoRMS data.

2.18.2 Semantic Interoperability

2.18.2.1 Metadata format(s)

WoRMS does not provide metadata (as “metadata” can be understood in the context of FAIR EASE).

2.18.2.2 Data format(s)

The WoRMS web services provide output in JSON. The HTML representation of any WoRMS name (the taxonomic and associated information for any AphiaID) are also provided as an RDF representation. Work has begun on providing this as Linked Open Data, and this will be ready for testing in 2024.

2.18.2.3 Use of controlled vocabularies - which, where, how

WoRMS uses Marine Regions IDs for location information, and the AphiaIDs themselves can be considered a controlled vocabulary.

2.18.3 Available web services for metadata access

None (see comment above).

2.18.4 Available web services for data access

2.18.4.1 Data access policy - if yes, which and how deployed

CC BY licence for all WoRMS data - see <https://www.marinespecies.org/about.php#terms> for the Terms of Use.

2.18.4.2 Any web services and API's - URLs, function, how to operate

For more information on the web services, see:

- <https://www.marinespecies.org/aphia.php?p=webservice>

Documentation for the REST services:

- <https://www.marinespecies.org/rest/>

For SOAP services, see

- Web Service Definition Language File:
 - <https://www.marinespecies.org/aphia.php?p=soap&wsdl=1>
- Brief description:
 - <https://www.marinespecies.org/aphia.php?p=soap>

2.19 European Marine Omics Biodiversity Observation Network

Content to be filled in over the coming months.

2.19.1 Data Discovery and Access Component

2.19.1.1 Name & Web Address

2.19.1.2 Types and number of data sets and/or data products

2.19.1.3 Discovery and Access mechanism and functionality

2.19.2 Semantic Interoperability

2.19.2.1 Metadata format(s)

2.19.2.2 Data format(s)

2.19.2.3 Use of controlled vocabularies - which, where, how

2.19.3 Available web services for metadata access

2.19.4 Available web services for data access

2.19.4.1 Data access policy - if yes, which and how deployed

2.19.4.2 Any web services and API's - URLs, function, how to operate

3 Other (Data) Infrastructures

3.1 European Gliding Observatories (EGO)

EGO: <https://www.ego-network.org/dokuwiki/doku.php>

The EGO initiative was launched by several teams of oceanographers, interested in developing the use of [gliders](#) for ocean observations and helped to set up a strong glider community. The glider is still a relatively new platform in oceanography, but one which has great [potential](#) thanks to its smart [design](#).

Gliders were developed in the 90's to carry out in-situ observations of the upper 1km of the ocean, filling the gaps left by the existing observing systems. Gliders look like small autonomous robotic underwater vehicles that use an engine to change their buoyancy. While gliding from the surface to about 1000 metres, gliders provide real-time physical and biogeochemical data along their transit. They observe temperature, salinity, pressure, biogeochemical data or acoustic data.

Our goal is to share the efforts needed by glider data collection as a community, and support the dissemination of glider data in global databases (like [Coriolis Data Center](#)) in real-time and delayed mode for a wider community.

3.2 Marine Mammals Exploring the Oceans Pole to Pole (MEOP)

MEOP: <https://www.meop.net/database/meop-databases/>

The MEOP data portal distributes currently three different databases:

- the [MEOP-CTD database](#): quality-controlled CTD profiles
- the [MEOP-SMS database](#): submesoscale-resolving high density CTD data
- the [MEOP-TDR database](#): high spatial density temperature/light data

To download the public data, you have to fill in a simple form following [this link](#), asking for your name, email, who you are and what you plan to do with the data. You will then access a page with the links to the data. The MEOP-CTD, MEOP-SMS and MEOP-TDR databases can be accessed through a zip file or through a read-only Dropbox folder. Note that the MEOP-CTD database can also be accessed from [SEANOE](#) (doi: 10.17882/45461). Please do not forget to [cite us](#) in your publications, as it is the best way to support us.

3.3 Glodap (Global Ocean Data Analysis Project)

Glodap: <https://www.glodap.info/>

The Global Ocean Data Analysis Project (GLODAP) is a synthesis activity for ocean surface to bottom biogeochemical data collected through chemical analysis of water samples. Data for 13 core variables (salinity, oxygen, phosphate, nitrate, silicate, dissolved inorganic carbon, total alkalinity, pH, CFC-11, CFC-12, CFC-113, CCl₄ and SF₆) are subjected to primary and secondary quality control to identify outliers and correct for measurement biases.

GLODAP is publicly available, discoverable, and citable. GLODAP enables quantification of the ocean carbon sink, ocean acidification and evaluation of ocean biogeochemical models. You can view GLODAP data using the interactive [Digital Earth Viewer](#).

3.4 Global Runoff Data Centre (GRDC)

GRDC: https://www.bafg.de/GRDC/EN/Home/homepage_node.html

The GRDC is an international archive of data up to 200 years old, and fosters multinational and global long-term hydrological studies. Originally established three decades ago, the aim of the GRDC is to help earth scientists analyse global climate trends and assess environmental impacts and risks. Operating under the auspices of WMO the database of quality controlled “historical” mean daily and monthly discharge data grows steadily and currently comprises river discharge data of well over 10,000 stations from 159 countries.

The exchange of hydrological data and information on a global scale is the principal reason for operating GRDC. National Hydrological and Hydro-Meteorological Services as primary providers of river discharge data and associated metadata and the scientific research community utilising this unique data collection are partners connected by GRDC. Positioned as a facilitator for exchanges between data providers and data users, the GRDC has become a focal point for international cooperation. Researchers, universities and other organisations make use of the facilities available for research programmes and projects.

GRDC data is available via webODV:

- <https://explore.webodv.awi.de/rivers/discharge/grdc/daily/>

3.5 Global Historical Climatology Network - Daily (GHCN-Daily)

GHCN: <http://ec2-52-38-26-42.us-west-2.compute.amazonaws.com:8080/dataset/noaa-ncdc-c00861>

The Global Historical Climatology Network - Daily (GHCN-Daily) dataset integrates daily climate observations from approximately 30 different data sources. Version 3 contains station-based measurements from well over 90,000 land-based stations worldwide, about two thirds of which are for precipitation measurement only. Other meteorological elements include, but are not limited to, daily maximum and minimum temperature, temperature at the time of observation, snowfall and snow depth. Over 25,000 stations are regularly updated with observations from within roughly the last month. The dataset is also routinely reconstructed (usually every week) from its roughly 30 data sources to ensure that GHCN-Daily is generally in sync with its growing list of constituent sources. During this process, quality assurance checks are applied to the full dataset. Where possible, GHCN-Daily station data are also updated daily from a variety of data streams. Station values for each daily update also undergo a suite of quality checks.

GHCN data is available via webODV:

- <https://explore.webodv.awi.de/atmosphere/meteorology/ghcn/daily/>

3.6 Marine Regions gazetteer

Marine regions gazetteer: <https://www.marineregions.org/about.php>

The purpose of Marine Regions is to create a standard, relational list of geographic names, coupled with information and maps of the geographic location of these features. This will improve access and clarity of the different geographic, marine names such as seas, sandbanks, ridges and bays and display univocally the boundaries of marine biogeographic or managerial marine areas.

Marine Regions is an integration of the VLIMAR Gazetteer and the VLIZ Maritime Boundaries Geodatabase. The VLIMAR Gazetteer is a database with geographic, mainly marine names such as seas, sandbanks, seamounts, ridges, bays or even standard sampling stations used in marine research. The geographic cover of the VLIMAR gazetteer is global but initially focused on the Belgian Continental Shelf and the Scheldt Estuary and the Southern Bight of the North Sea. Gradually more regional and global geographic information was added to VLIMAR and combining this information with the Maritime Boundaries database, representing the Exclusive Economic Zone (EEZ) of the world, led to the creation of marineregions.org.

3.7 BiG-FAM database

The BiG-FAM database : <https://bigfam.bioinformatics.nl/home>

The Big-FAM database provides an explorable "atlas" of microbial secondary metabolic diversity to browse and search biosynthetic diversity across taxa. It is based on the Biosynthetic Gene Cluster Family (GCF) database which is an online repository for homologous groups of biosynthetic gene clusters (BGCs). It contains GCFs calculated by the BiG-SLiCE algorithm from over 1.2 million genomes, and allows users to easily search and browse them to analyse patterns of biosynthetic diversity across taxa.

The last database version (1.0.0) is based on a GCF clustering of 1,225,071 BGCs taken from multiple publicly available sources. 29,955 GCF models, each representing distinct protein domain and sequence features shared by the BGCs are represented in the database.

Users can browse the GCFs database and filtered the result by type of dataset (e.g. `isolate_archaeal`, `isolate_bacterial_complete`, `mag_ocean`, etc.) or by BGC class (Alkaloid, Saccharide, etc.) to find out datasets of interest. They can also browse the BGCs with the same filtering criteria. Finally, a query interface is available where users can insert an antiSMASH job.

3.8 AERONET Aerosol Network

AERONET Aerosol Network: <https://aeronet.gsfc.nasa.gov/>

The AERONET (AErosol RObotic NETwork) program is a federation of ground-based remote sensing aerosol networks established by [NASA](#) and [PHOTONS](#) and is greatly expanded by networks (e.g., [RIMA](#), [AeroSpan](#), [AEROCAN](#), [NEON](#), and CARSNET) and [collaborators](#) from national agencies, institutes, universities, individual scientists, and partners. For more than 25 years, the project has provided long-term, continuous, and readily accessible public domain databases of aerosol optical, microphysical and radiative properties for aerosol research and characterization, validation of satellite retrievals, and synergism with other databases. The network imposes standardisation of [instruments](#), [calibration](#), [processing](#) and [distribution](#).

4 Conclusions

The Interdisciplinary Data Discovery and Access Service of the FAIR-EASE project aims at including the 18 (data) infrastructures that have been described in Chapter 4. The infrastructures have been analysed based on their current data discovery and access mechanisms. In the table below the most important findings regarding this initial analysis are listed.

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
Copernicus Marine Service (CMEMS)	JSON based REST API , OGC CSW AP ISO profile	ISO 19139 based MyOcean profile	P22 , MyOcean controlled vocabularies (temporal scale, discipline, area of benefit, reference geographical areas, processing level, model assimilation, feature type)	HTTPs - MOTU API , HTTPs - WMS API , HTTPs - WCS API , HTTPs - ERDDAP API , HTTPs - OPeNDAP API , FTP - File Transfer Protocol , REST API
EMODnet, Chemistry, Physics, Bathymetry, Biology	OGC CSW AP ISO catalogues	ISO 19139 based profile	P01 , P02 , P06 , P24 , P07 , L05 , L22 , P35	WMS, WFS, WCS, REST API, THREDDS, ERDDAP (endpoints)
Euro-Argo	OpenAPI REST API endpoints (fleet monitoring , dataselection)	The netCDF file includes Euro-Argo metadata elements in CF-netCDF compliant format	http://www.argodatamgt.org/Documentation/Argo-vocabulary-server	ERDDAP , FTP , Collections(DOI , GDAC)
SeaDataNet <ul style="list-style-type: none"> Common Data Index (CDI) Data Products & Sextant web catalogue 	SPARQL service for CDI metadata, OGC WMS service for locations of all data sets, OGC WFS service for metadata features of all data sets Sextant OGC CSW catalogue	SeaDataNet CDI metadata profile is reported to be an INSPIRE compliant profile of ISO 19115, its encoding is made according to XML (ISO 19139).	P01 , P02 , P06 , P24 , P07 , L05 , L22	
GEBCO	None present (except for the direct download of the full dataset as NetCDF)	The netCDF file includes metadata elements in CF-netCDF compliant format (1.6)	P01 , P06	Direct download of all the GEBCO grid as NetCDF or GeoTIFF,

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
				the current GEBCO grid and TID grid (and previous versions of the grid) are as well available as WMS
NASA Earth Observation Data	There is an Earthdata Developer Portal with information of relevant APIs that can be used to search and access Earth science data. Probably CMR Search API is the preferred option.	ATOM CSV DIF-9 DIF-10 ECHO 10 ISO-MENDS ISO-SMAP JSON UMM JSON KML Open Data STAC XML Reference	Global Change Master Directory (GCMD) keywords, which contain a set of Earth Science vocabularies	There is an Earthdata Developer Portal with information of relevant APIs that can be used to search and access Earth science data.
Copernicus Land Monitoring Service	Global catalogue CSW Pan-European, Local, Imagery and reference data catalogue CSW	The metadata format is reported to be an INSPIRE compliant profile of ISO 19115, its encoding is made according to XML (ISO 19139).	P22 , P07	None seems to be present for complete machine to machine interaction
European Space Agency <ul style="list-style-type: none"> Copernicus Open Access Hub EUMETSAT 	The following Open Search URI addresses the resource /Products exposed by the Open Search Service ----- OGC CSW endpoint REST API endpoint The Data Store OpenSearch API The Data Store Browse REST API	ISO 19139 metadata are only available upon data product download. ----- --- EUMETSAT metadata profile of ISO 19115-2 metadata records can be retrieved from the CSW		To download a full products OData service can be used, with the following syntax: <ServiceRootUri>/Products('Id')/\$value Example: https://scihub.copernicus.eu/dhus/odata/v1/Products('2b17b57d-fff4-4645-b539-91f305c27c69')/\$value ----- ----

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
	Available Data Store collections are also listed here			<p>The Data Store Download REST API</p> <p>Products available through the EUMETview, WMS, WFS, WCS interfaces are as well available for those products:</p> <ul style="list-style-type: none"> • WMS • WFS • WCS
Climate Data Store	REST API at https://cds.climate.copernicus.eu/solr/drupal/select?	<p>Returned metadata is JSON encoded, basic elements are present.</p> <p>The data sets are in GRIB1, GRIB2 or NetCDF4 format. The GRIB formats use the WMO standards and the NetCDF format follows the CF-1.4, CF-1.5 or CF-1.6 convention.</p>		Climate Data Store (CDS) Application Program Interface (API) available for resources of type “dataset”
<p>NOAA</p> <ul style="list-style-type: none"> • Physical Sciences Laboratory • National Centers for Environmental Information 	<p>THREDDS catalogue</p> <p>-----</p> <p>The CSW catalogue of the NCEI Geoportal</p>	<p>Each data set has a corresponding metadata file in XML using the ISO19115 metadata standard. There is also a NCML metadata representation available and an UDDC evaluation of how well the metadata contained in the data conforms to the NACDD (see for example this page).</p> <p>-----</p>		<p>OpenDAP, see this page for more information</p> <p>HTTPServer, HTTP file download</p> <p>WCS, which supports the access to geospatial data as coverages</p> <p>WMS, which supports the access to georegistered map images from geoscience data sets</p>

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
		The data sets have a corresponding metadata file in XML using the ISO19115-2 metadata standard.		NETcdfSubset, which is a web service for subsetting CDM scientific grid datasets. Use FTP, see this page for more information ----- ----There is a Search API available that provides an endpoint for the NOAA OneStop data discovery. The documentation behind the API can be found here and here .
European Soil Data Centre	There are no available web services for metadata access.	Each dataset is described by a detailed metadata page. Metadata fields are not fully harmonised between datasets.	There is no controlled vocabulary used for describing the variables stored in the data.	The data download has to be done from each dataset webpage. For dataset requiring a prior registration through a fill-in form, the download url is sent by email.
Joint Research Centre Data Catalogue	CKAN action API , ODCAT action API	DCAT-AP (DCAT application profile for data portals in Europe)		Direct download
European Environment Agency	OGC CSW AP ISO	EEA metadata profile , GeoTIFF, Ascii, SHP, Excel, Access, ...	P22 , EEA topics, INSPIRE spatial scope	Direct download, WMS
National Center for Biotechnology Information (NCBI)	To get machine readable information the OBO Foundry ontology has to be	Ontology Web Language (owl) and RDF	Only for the OBO foundry version of the NCBI, it is an ontology that uses owl and RDF. The	Another service was found outside of NCBI that offers a json output using REST:

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
	used using this link with a taxonid: https://ontobee.org/ontology/rdf/NCBITaxon?iri=http://purl.obolibrary.org/obo/NCBITaxon_ID		taxonomy browser is in csv, ascii and html.	https://rest.ensembl.org/taxonomy/id/1386?content-type=application/javascript
Global Biodiversity Information Facility (GBIF)	GBIF Registry API	Ecological Metadata Language (EML), Darwin Core Archive	The DwC-A terms are used in the DwC-A files, and all “measurements and facts” are linked to controlled vocabularies: which ones are used depends on the data and the data creator; those from the ENVO ontology, and BODC, WoRMS, and Marine Regions for the ocean-linked data, are certainly used, as are MiXS and NCBI (and other omics) IDs for DNA-based data.	Dataset direct download, GBIF Occurrence API
Ocean Biodiversity Information System (OBIS)	OBIS REST API	EML, RTF		Dataset direct link, granules download through OBIS REST API
European Nucleotide Archive (ENA)	ENA Portal API	TSV, JSON, XML based, EMBL Flatfile, FASTA		FTP
World Register of Marine Species	For more information on the web services, see: https://www.marinespecies.org/aphia.php?p=webservice Documentation for the REST services:	WoRMS does not provide metadata	WoRMS uses Marine Regions IDs for location information, and the AphiaIDs themselves can be considered a controlled vocabulary.	

(Data) Infrastructure	Discovery web services	(Meta) data standard/format	Vocabularies	Access web service
	<p>https://www.marinespecies.org/rest/</p> <p>For SOAP services, see Web Service Definition Language File: https://www.marinespecies.org/aphia.php?p=soap&wsdl=1</p>			