ODYSSEA

Operating a network of integrated observatory systems in the Mediterranean Sea

Project Deliverable Report

Deliverable Number: 3.3

Deliverable Title: Data Management Plan for Collected Data

Author(s): Adélio Silva, Carlos Figueiredo, Inês Pedro, João Rodrigues, Pedro Galvão, Paulo Leitão, Laudy Claire

Work Package Number: 3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727277

ODYSSEA Project Information	
Project full title	ODYSSEA: Operating a network of integrated observatory systems in the Mediterranean Sea
Project acronym	ODYSSEA
Grant agreement number	727277
Project coordinator	Georgios Sylaios, DUTH
Project start date and duration	1st June 2017, 54 months
Project website	http://ODYSSEAplatform.eu/

Deliverable Information	
Work package number	3
Work package title	Designing the Platform
Deliverable number	3.3
Deliverable title	Data Management Plan for Collected Data
Description	
Lead beneficiary	Hidromod
Lead Author(s)	Adelio Silva
Contributor(s)	Carlos Figueiredo, Inês Pedro, João Rodrigues, Pedro Galvão, Paulo Leitão, Georgios Sylaios
Revision number	V0.5
Revision Date	April 30, 2018
Status (Final (F), Draft (D), Revised Draft (RV))	F

Dissemination level (Public (PU), Restricted to other program participants (PP), Restricted to a group specified by the consortium (RE), Confidential for consortium members only (CO))	Public

Document History			
Revision	**Date**	**Modification**	**Author**
0.1	22/10/2018	Initial draft starting from D3.2	Adélio Silva
0.2	05/11/2018	Provisional version	Adélio Silva
0.3	25/11/2018	Near Final version	Georgios Sylaios
0.4	25/11/2018	Document template modification	Nikolaos Kokkos
0.5	28/11/2018	Review and editing	Bracha Ehrman
0.6	30/11/2018	Final editing	Georgios Sylaios

Approvals				
	Name	**Organisation**	**Date**	**Signature (initials)**
Coordinator	Georgios Sylaios	DUTH		GSylaios
WP Leaders	Adélio Silva	HIDROMOD, Lda	05/11/2018	A.S.

Table of Contents

Table of Figures

1. Executive Summary

ODYSSEA intends to develop, operate and demonstrate an interoperable and cost-effective platform that fully integrates networks of observing and forecasting systems across the Mediterranean basin, addressing both the open sea and the coastal zone.

The platform is prepared to deliver a set of services focused on different coastal user's needs (navigation safety, ports operations, water pollution prevention and response, eutrophication risks, search and rescue missions, etc.) enabling the exploitation of the added value of integrated Earth Observation (EO) technologies (satellite, airborne and ground based), Copernicus Marine Service and ICT to deliver customized and ready to use information. These services will provide an easy way to access in-situ data, local high-resolution forecasts and products and services (e.g. meteo-oceanographic conditions at specific locations, identification of optimum or critical working windows, support to sea pollution response actions, etc.) for a broad range of different users and stakeholders.

Taking in consideration that this platform will gather a large number of diverse data sets (from existing networks and platforms and the ODYSSEA-produced data consisting of results from operational numerical models, data from in-situ sensors and remotely sensed data), the issues of data management and data quality control assume a central concern. One goal of the platform is to ensure that data from different and diverse data providers are readily accessible and useable to the wider oceanographic community. To achieve that, the strategy is to move towards an integrated data system within ODYSSEA that harmonizes work flows, data processing and distribution across different systems.

The value of standards is clearly demonstrable. In oceanography, there have been many discussions for processing data and information. Many useful ideas have been developed and put into practice, but there have been few successful attempts to develop and implement international standards in managing data.

This document intends to provide an overview of the best practices concerning these aspects and define the guidelines to be followed in themes such as catalogues, metadata, data vocabulary, data standards and data quality control procedures. This implies taking actions at different levels:

- Adopt proper data management procedures to implement metadata, provide an integrated access to data in order to facilitate integration into existing systems and assure the adoption of proper data quality control.
- Enable integration of more data, improve the enhancement of the services (viewing, downloading, traceability and monitoring) to users and providers, facilitate the discovery of data through a catalogue based on ISO standards, provide OGC services (SOS, WMS, WFS, etc.) to facilitate development; and ensure the visibility of existing data and the identification of gaps.

As ODYSSEA is an EU supported Mediterranean-focused platform, the data management plan will tune in all specific aspects of ocean data management within the European context, such as existing networks of data and requirements of European industry and other end-users. This deliverable will be the final and updated version of the Data Management Plan (DMP) for the ODYSSEA Platform, including strategies for improving data management, data privacy issues and data quality control procedures. In this updated

document, problems faced during the implementation process, barriers and lessons learnt will be discussed. It will further elaborate specific relevant aspects for ODYSSEA namely the "new" data series approach using SOS, further integration into the ODYSSEA Platform among others.

2. Introduction

ODYSSEA aims to provide a set of services focused on different coastal users' needs (navigation safety, ports operations, water pollution prevention and response, eutrophication risks, search and rescue missions, etc.) allowing to exploit the added value of integrated Earth Observation (EO) technologies (satellite, airborne and ground based), Copernicus Marine Service and ICT to deliver customized and ready to use information. These services will provide an easy way to get in-situ data, local high-resolution forecasts and products and services (e.g. meteo-oceanographic conditions at specific locations, identification of optimum or critical working windows, support to sea pollution response actions, etc.) to a broad range of different users.

This report describes the strategies to be implemented to improve data management, data privacy and data quality control of ODYSSEA services. The strategy has four main components:

- Catalogues, vocabulary and metadata;
- Data integration and fusion;
- Data quality control;
- Data privacy policy.

The issue of metadata, vocabulary and catalogues is of prime importance to assure the interoperability and easy discovery of data. A proper data management plan following widely accepted standards also contributes to the reduction in the duplication of efforts among agencies. Likewise, the plan is to improve the quality and reduce the costs related to geospatial data processing, thus making oceanographic data more accessible to the broader public while helping to establish key partnerships to increase data availability. Aiming to contribute to these objectives, ODYSSEA will adopt the procedures already proposed by the most relevant EU initiatives such as CMEMS, EMODNet and SeaDataNet, especially the standards in relation to vocabularies, metadata and data formats. In practice the gridded data sets addressing either dynamic data sets (similar to CMEMS) or static data sets (similar to EMODnet) will follow procedures similar to the ones adopted by these two services. Regarding the time series data, SeaDataNet procedures will represent the main guidelines and NetCDF-CF format will be the standard to be adopted. However, **ODYSSEA will go one step further and will use these NetCDF files to feed an SOS service, supported by North 52 software to assure the interface with the users of the platform.**

The capability of serving time-series through a standard protocol, such as SOS, will represent a step forward from the existing services although, as a pioneer, it is foreseen that ODYSSEA will be required to overcome some barriers. The service has been tested in the ODYSSEA platform V0 Edition and it is one of the subjects to be discussed and showcased in this updated/final version of the ODYSSEA DMP.

The data integration and fusion policies to be adopted in ODYSSEA are also relevant issues of the project. Data integration and fusion deals with the best strategies to adopt when it comes to merge datasets obtained from different data sources, building the best available datasets or fusing different data sources to produce aggregated data (i.e., secondary parameters and indicators). Although not easy, addressing this issue properly may represent a valuable contribution to improve data accuracy and robustness of models'

initial and boundary conditions and provide the users with comprehensive data that merge different data sets based on reliable criteria.

The data quality control either related with the quality of observed in-situ data (e.g. tidal gauges, wave buoys, weather stations, etc.) or the modelled forecasts is another relevant aspect that will be addressed by ODYSSEA's DMP. In the case of locally acquired data, automatic procedures will run regularly to detect and remove anomalous values from observed datasets. In the case of the models, the results will be automatically compared with observations (e.g., buoys and CMEMS grid observation products) and the statistical analysis will be provided on a daily basis to the end users.

Regarding data privacy (data protection and the rights of platform end-users, customers and business contacts), it is clear that ODYSSEA will respect personal data under the General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679) which will substitute Directive 95/46/EC on May 25, 2018. 'Personal data' means any information, private or professional, which relates or can be related to an identified or identifiable natural person (for the full definition, see Article 2(a) of EU Directive 95/46/EC).

In the following paragraphs a more detailed overview, both of the "state of the art" and the procedures to be adopted in ODYSSEA, will be provided.

3. Ocean Data Management: the European context

Delivery of data to users requires common data storage and transfer formats, which interact with other standards (Vocabularies, data quality control). Several initiatives exist within Europe for ocean data management, which are now coordinated under the umbrella of EuroGOOS. EuroGOOS is the network committed to develop and advance the operational oceanography capacity of Europe, within the context of the intergovernmental Global Ocean Observing System (GOOS). The scope of EuroGOOS is wide and its needs are partially addressed by the on-going development within Copernicus, SeaDataNet and other EU initiatives.

Therefore, to improve the quantity, quality and accessibility of marine information, to support decision making and to open up new economic opportunities in the marine and maritime sectors of Europe for the benefit of European citizens and the global community, it was agreed at the annual EuroGOOS meeting in 2010 that it is essential to meet the following needs (AtlantOS, 2016):

- Provide easy access to data through standard generic tools; where “easy” means the direct use of data without concerns on data quality and processing and that adequate metadata are available to describe how the data were processed by the data provider.
- Combine in situ-observation data with other information (e.g., satellite images or model outputs) to derive new products, build new services or enable better-informed decision-making.

The ocean data management and exchange process within EuroGOOS intends to reduce the duplication of efforts among agencies as well as to improve the quality and reduce the costs related to the geospatial information, thus making oceanographic data more accessible to the public and helping to establish key partnerships to increase data availability. In addition, the EuroGOOS data management system intends to deliver a system that will meet European needs, in terms of standards while respecting the structures of the contributing organizations.

The structure will include:

- Observation data providers, which can be operational agencies, marine research centres, universities, national oceanographic data centres and satellite data centres.
- Integrators of marine data, such as the Copernicus in-situ data thematic centre (for access to near real-time data acquired by continuous, automatic and permanent observation networks) or the SeaDataNet infrastructure (for quality controlled, long-term time series acquired by all ocean observation initiatives, missions, or experiments), ICES and EurOBIS for biodiversity observations, and the new European Marine Observation and Data Network (EMODnet) portals. The integrators that will support both data providers, willing to share their observation data, and users requesting access to oceanographic data (historic, real-time and forecasts). Integrators develop new services to facilitate data access and increase the use of both existing and new observational data.
- Links with international and cross-disciplinary initiatives, such as GEOSS (Global Earth Observation System of Systems), both for technical solutions to improve harmonization in an interdisciplinary global context.

3.1. Towards an integrated EU data system

ODYSSEA aims to contribute to improving data availability for end-users and stakeholders across the Mediterranean basin, addressing both the open sea and the coastal zone. One goal is to ensure that data from different and diverse in-situ observing networks and forecasting models are readily accessible and useable. **To achieve this, the strategy is to move towards an integrated data system that harmonizes work flows, processes data according to existing standards and disseminates data produced by the in-situ observing and modelling network system, while integrating in-situ observations into existing European and international data infrastructures** (the so called "Integrators"). Such Integrators include: the Copernicus INS TAC, SeaDataNet NODCs, EMODnet, EurOBIS, and GEOSS.

The targeted integrated system deals with data management challenges that must be met to provide efficient and reliable data service to users. These include:

- Common quality control for heterogeneous and near real time data;
- Standardization of mandatory metadata for efficient data exchange;
- Interoperability of Network and Integrator data management systems.

3.2. Industry requirements

Presently, there is a need to change the way marine observatories and public data-sharing initiatives engage with industry and users. The Columbus project (funded by the EU under H2020 which ended this year) proposes a set of recommendations designed to overcome some of the most important gaps and barriers still faced by private data users. Taken together, they represent the basic components of a strategy to open significant opportunities for the maritime industry to both benefit from and engage with public marine data initiatives. This can ensure the optimum return of public investments in the marine data sector, notably in support of meeting key EU policy goals under the Blue Growth Strategy, the Marine Strategy Framework Directive and the Maritime Spatial Planning Directive. Some barriers require further analysis and discussion, but there are already many actions that can be undertaken to improve the situation on the short and medium term (Columbus, 2017):

- Industry representatives should be included in the governance and take part in the entire cycle of decision making, development and operation of marine observation and data-sharing initiatives.
- There is a need for marine data-sharing initiatives to take a more pro-active approach and move out of the comfort zone of the traditional oceanographic marine monitoring and observing communities. This involves, among others, developing a more "service-oriented approach", learning new communication skills and language, being present and more visible in fora that attract industry and to exploit creative technologies.
- Data, products and services offered by marine observation and data initiatives should be presented in a user-friendly, attractive and intuitive way which is adapted to the target users. If users from different communities or sectors are targeted, options to adjust the interface depending on the visitor should be considered.

- Clear, succinct and open communication is critical: it should be instantly clear for industry what data, products and services are offered and what may be made available in the future. Equally important is to provide information on what is not available, and the limitations of the resources offered.
- More efforts should be made to build upon early achievements and successes: presenting use case examples that can trigger interest where there may previously have been none.
- There is a significant role for maritime clusters in connecting marine data initiatives with industry and vice versa. Maritime clusters are an important bridge between private and public sector as they deal with both and have a good understanding of their culture, language, needs and concerns.
- At European level there is a need for defragmentation of the plethora of marine observation and data and information sharing initiatives, as well as the online data portals. In the longer term, there is a need for a joint roadmap, agreed by the responsible coordinating and funding bodies including at the European Commission level, to set out the strategic framework.
- Dedicated data-sharing policies to incentivise the private sector and address their specific needs should be developed. Ways forward could include: stating clearly the added-value or benefits of sharing data, moratorium on commercially sensitive data, provision of services in return for data which could support in-house data management, the development of a data-sharing 'green label' in recognition of corporate social responsibility. It is clear that implementation of the recommendations will require increased commitment and investment of time and resources, both from industry and from marine observation and data initiatives, but should provide both with significant returns over time

3.3. The ODYSSEA approach

The procedures to follow in ODYSSEA regarding this issue of the data management will preferentially follow the examples from CMEMS, EMODNet or SeaDataNet. In practice two major data types will be addressed: the gridded data produced by ODYSSEA models and the time-series data reported by the ODYSSEA static systems. Similarly, for the spatio-temporal data produced via sensors integrated into the ODYSSEA gliders, the SeaDataNet netCDF data standards for profiling along trajectories will be adopted.

The gridded data may address dynamic data sets (similar to CMEMS) or static data sets (similar to EMODnet). In both cases the procedures to follow will be similar to the ones adopted by these two services.

Regarding the time series data, SeaDataNet procedures will represent the main guidelines and netCDF-CF format will be the standard to be adopted. However, **ODYSSEA will go one step forward and will use these netCDF files to feed an SOS service, supported by North 52 software to assure the interface with the users.**

4. Data quality control

The issue of data quality control will be addressed following the state-of-the-art recommendations of different projects such as SeaDataNet or AtlantOS. SeaDataNet produced a comprehensive document presenting a set of guidelines to be followed in marine data quality control. According to this document, quoted below, data quality control essentially and simply has the following objective: "*To ensure the data consistency within a single data set and within a collection of data sets and to ensure that the quality and errors of the data are apparent to the user who has sufficient information to assess its suitability for a task*". If done well, quality control brings about several key advantages (SeaDataNet, 2010):

- ***Maintaining Common Standards****: There is a minimum level to which all oceanographic data should be quality controlled. There is little point banking data just because they have been collected; the data must be qualified by additional information concerning methods of measurement and subsequent data processing to be of use to potential users. Standards need to be imposed on the quality and long-term value of the data that are accepted (Rickards, 1989). If there are guidelines available to this end, the end result is that data are at least maintained to this degree, keeping common standards to a higher level.*
- ***Acquiring Consistency****: Data within data centres should be as consistent to each other as possible. This makes the data more accessible to the external user. Searches for data sets are more successful as users are able to identify the specific data they require quickly, even if the origins of the data are very different on a national or even international level.*
- ***Ensuring Reliability****: Data centres, like other organisations, build reputations based on the quality of the services they provide. To serve a purpose to the research community and others their data must be reliable, and this can be better achieved if the data have been quality controlled to a 'universal' standard. Many national and international programmes or projects carry out investigations across a broad field of marine science which require complex information on the marine environment. Many large-scale projects are also carried out under commercial control such as those involved with oil and gas and fishing industries. Significant decisions are made, and theories formed, on the assumption that data are reliable and compatible, even when they come from many different sources.*

ODYSSEA services data flux will be managed automatically by the ODYSSEA platform. The data quality control will start by the execution of automatic procedures (independently of the adoption of more complex procedures). The data quality control methodology will focus on in situ observations and modelled forecasts and it will be addressed from two perspectives: the data **Quality Assurance** and the **Quality Control**.

Quality Assurance (QA) is a set of review and audit procedures implemented by personnel or an organization (ideally) not involved with normal project activities to monitor and evaluate the project to maximize the probability that minimum standards of quality are being attained. With regard to data, QA is a system to assure that the data generated is of known quality and well-described data production

procedures are being followed. This assurance relies heavily on the documentation of processes, procedures, capabilities, and monitoring. Reviews verify that data quality objectives are being met within the given constraints. QA is inherently a human-in-the-loop effort and substantial documentation must accompany any QA action. QA procedures may result in corrections to data. Such corrections shall occur only upon authorized human intervention (e.g., marine operator, product scientist, quality analyst, principal investigator) and the corrections may either be applied in bulk (i.e., all data from an instrument during a deployment period) or to selective data points. The application of QA corrections will automatically result in the reflagging of data as 'corrected'.

Quality Control (QC) is a process of routine technical operations, to measure, annotate (i.e., flag) and control the quality of the data being produced. These operations may include spike checks, out-of-range checks, missing data checks, as well as others. QC is designed to:

- Provide routine and consistent checks to ensure data integrity, correctness, and completeness;
- Identify and address possible errors and omissions;
- Document all QC activities.

QC operations include automated checks on data acquisition and calculations by the use of approved standardized procedures. Higher-tier QC activities can include additional technical review and correction of the data by human inspection. QC procedures are important for:

- Detecting missing mandatory information;
- Detecting errors made during the transfer or reformatting;
- Detecting duplicates;
- Detecting remaining outliers (spikes, out of scale data, vertical instabilities, etc);
- Attaching a quality flag to each numerical value to indicate the corrected observed data points.

A guideline of recommended QC procedures has been compiled by the SeaDataNet project after reviewing NODC schemes and other known schemes (e.g. WGMDM guidelines, World Ocean Database, GTSPP, Argo, WOCE, QARTOD, ESEAS, SIMORC, etc.). The guideline at present follows the QC methods proposed by SeaDataNet for CTD (temperature and salinity profiles), current meter data (including ADCP), wave data and sea level data. SeaDataNet is also developing efforts for extending the guideline with QC methods for surface underway data, nutrients, geophysical data and biological data.

ANNEX I provides a detailed description of the implementation process procedure to be followed for QA/QC in ODYSSEA.

4.1. Quality Control Flags

According to EuroGOOS (2016), an extensive use of flags to indicate the data quality is recommended, since the end user will select data based on quality control flags, amongst other criteria. These flags should always be included in any data transfer (e.g., from ODYSSEA Observatories to the central ODYSSEA platform) maintaining standards and ensuring data consistency and reliability (*see Table 1*). The same flag scale is also recommended by SeaDataNet.

TABLE 1: QUALITY FLAG SCALE (REPRODUCED FROM EUROGOOS, 2016).

Code	Definition
0	No QC was performed
1	Good data
2	Probably good data
3	Bad data that are potentially correctable
4	Bad data
5	Value changed
6	Bellow detection limit
7	In excess of quoted value
8	Interpolated value
9	Missing value
A	Incomplete information

- Data with QC flag = 0 should not be used without a quality control made by the user.
- Data with QC flag different from 1 on either position or date should not be used without additional control from the user.
- If date and position QC flag = 1 only measurements with QC flag = 1 can be used safely without further analyses
- if QC flag = 4 then the measurements should be rejected
- if QC flag = 2 the data may be good for some applications, but the user should verify this
- if QC flag = 3 the data are not usable, but the data centre may be able to correct them in a delayed mode

4.2. In-situ observations quality control

The quality control of observations can be done in two phases. During the download of in-situ observations automatic checks should be done such as those proposed by SeaDataNet (2010) (e.g. global range test, date and time). After quality control, only the valid data is stored in the database. In the second phase, a tool may be run periodically to perform a scientific quality control check (SeaDataNet, 2010). This quality control aims to detect spikes, filter high frequency noise (e.g. moving average or P50), data with abnormal variability in time, etc. Specific tools will be running automatically for this purpose.

4.3. Forecasts quality control

The quality control of modelled forecasts can be done by comparing time-series forecasts with in-situ observations (e.g., wave buoys, tidal gauge, weather stations, etc.) through automatically-run algorithms.

Similarly, gridded data forecasts may be compared automatically with observations (e.g., CMEMS gridded data observations). As a result, several statistical parameters may be computed (e.g., correlation coefficient, bias, RMSE, skill, etc.) to assess the quality of forecasts.

QA/QC procedures will be followed as the data from local Observatory PCs reach the central ODYSSEA platform. An extensive analysis of ODYSSEA QA/QC procedures is provided in Section 7.3 "Quality Control and Data Processing Functionality" of Deliverable 2.3.

5. Data integration and fusion

5.1. Low level data integration and fusion

Adopting the best strategies for merging datasets obtained from different data sources, in order to build the best available datasets or fuse different data sources to produce aggregated data, indices and products, is not simple. A possible solution when we have different solutions with different resolutions for the same area is to fuse these data and offer a unique integrated dataset. Another option is to provide all datasets separately with an option of an integrated solution. No matter which solution is adopted, the final objective of the data integration and fusion is to contribute to improvement of data accuracy and robustness of models' initial and boundary conditions as well as to provide users with comprehensive data that merge different data sets based on reliable criteria.

For example, if a user is interested in operational wave data for a specific site and realizes that for the period of interest, there exist different time series from different wave buoys, he may be interested in getting a unique time series by merging different time series data and making them compatible. This process may require complex actions regarding the levels of accuracy of the different measuring devices, the measuring time rate and units, etc.

5.2. Semantic Information Integration and Fusion

Capacity for integration and fusion of semantic information will be provided through the ODYSSEA platform. Semantic information is composed of several pieces of information, potentially coming from different semantically rich information sources. The main use of this capacity is for semantic network enrichment and query.

The information processed is expressed through graphs of entities related with each other and contains semantic metadata. The fusion is adapted to the domain of application. This application domain is described through an ontology of the domain. The fusion process is also adapted to the quality of the information items, through the use of fusion heuristics.

The fusion heuristics integrate domain knowledge and user preferences. They are the intelligent part of the semantic fusion system. They are end-user defined functions used to express the confidence the users have in the information sources, as well as specific strategies that must be followed in order to fuse information coming from different sources.

The two main semantic information integration functionalities are:

- Insertion of new information into a semantic information network (Synthesis),
- Query for information in a semantic information network (Mining).

5.2.1. Information sources

Many valuable open information sources can be used and integrated in order to provide a large and always up to date overview of the ongoing situation in specific zones. For instance, we will use information provided by the Wikipedia[1] encyclopaedia, the Wikidata[2] information base and social networks such as Twitter. For scientific data sets, bases such as the EMODnet platform can be used to provide data on main ports activities, quality of bathing waters … etc.

Wikipedia pages

Wikipedia encyclopaedia is a wide collaborative and constantly up-to-date source of information. Integrating information provided by the Wikipedia community enables providing end-users with a very rich semantic source of information. Regarding domains such as tourism, locations and marine species, the information available on Wikipedia is particularly abundant.

Wikidata elements

Wikidata is another valuable source of semantic information. Contrary to the information stored on Wikipedia, Wikidata is a knowledge base thus the semantic aspect of information is contained in the information source. Concepts and instances are defined in Wikidata. Semantic relations among these objects are specified. Therefore, Wikidata is an open source of information of valuable importance.

EMODnet Human Activities

Among providers, EMODnet provides several data portal for marine data. Some of the portals, such as the Human Activities portal, provide information with a rich semantics and a high level of interpretation. This source may be of great interest in order to be integrated with other sources of information, providing different perspectives on the marine situation of the different zones of interest.

Twitter

Social media provides a wealth of crowd sourced data on easily observable physical phenomenon in settings which lack traditional methods of monitoring and can even prove to be more rapid and flexible. However, the challenge lies in mining actionable data from the millions of tweets posted every hour. Hence a collection/filtering algorithm will be written in order to collect tweets which are contextually and geographically pertinent.

The twitter API may be used to retrieve raw data that for further manipulation and processing in order to get useful information. The raw data supplied by a Twitter API call consists of JSON objects, which contain a large number of categories of information (an attribute followed by human readable text).

---

[1] *Wikipedia, the free encyclopedia : en.wikipedia.org*

[2] *Wikidata : wikidata.org*

5.2.2. Need for information quality enhancement algorithms

If most of the encyclopaedic sources of information are reliable and complete, social media information has a very low level quality. However, using social media is of importance if we want to involve citizens in the monitoring and protection of our environment. For example, in ODYSSEA we are building the so-called "models chain", where each model is run in a predetermined order and the data produced by one model are used as boundary conditions for the next. In this chain, several uncertainties may occur, for example the initiation of the oil spill model when an oil spill accident occurs. **Through the ODYSSEA semantic information analysis, and more specifically through algorithms searching and harvesting the Twitter for relevant information, a more instant response to a disastrous event might occur.** Thus, the Oil Spill Model could be initiated at the exact location as early as possible to the event of oil release to the marine environment. Similar applications might include extreme events on meteorologic/hydrgraphic conditions (e.g., storms), eutrophication/bloom incidents, jelly-fish outburst, etc.

One of the issues of using Twiter for instance, is that due to the increasing digital data privacy restrictions the twitter users have to actively consent to provide their exact geolocation when posting their tweets. Otherwise the alternative is using the self-reported location of the profile associated with a given tweet, but this is usually very generalized (e.g. Europe), and hence unfit for this purpose. Hence, the only reliable option for gleaning geographic information of the tweet is by using keywords within the body of the tweet (e.g. @laplaya, #Madrid, etc.).

To overcome this limitation and be able to use citizen information provided through Twitter, it is necessary to deeply analyse the meanings (semantics) of the texts in order to understand it. Regarding locations, for instance, it is required to make use of a Named Entity Extraction engine in order to extract the location of the events reported in the tweets, rather than the location of the phones that were used to tweet.

Furthermore, authors of tweets have very different levels of reliability regarding specific issues. Average citizens won't know the species of a specific jellyfish that that see on a beach for instance. They may not be able to report properly and with all characteristics, an oil spill they witness. More serious issues may be encountered with authors that spread rumours or even create false information. A possible action to overcome such limitations is to build lists of referenced Twitter accounts regarding each use case, based on the accounts that the ODYSSEA end-users trust. For instance, maritime governmental organisation accounts will be followed and analysed.

6. Data management

6.1. Providers code for data

Following the procedures adopted by AtlantOS, the Institutions providing data to ODYSSEA platform should be reported and acknowledged following the EDMO code recorded in the data file and the ODYSSEA platform catalogue. EDMO is the European Directory of Marine Organizations, developed under SeaDataNet, and it can be used to register any marine organization involved in the collection of datasets (operators, funders, data holders, etc.). It delivers a code for the organization to be included in the data or metadata leading to the harmonization of information (compared to free text) and the optimization of the datasets discovery. EDMO is coordinated by MARIS.

For EU Countries new entries are added by the National Data Centres (NODCs). Through ODIP (Ocean Data Interoperability Platform) cooperation, there is also a point of contact with the USA, Australia and some other non-EU countries. The rest of the world is managed by MARIS, which also moderates the first entrance in EDMO of new entries.

The request for a new entry in EDMO is sent to MARIS (current contact: Peter Thijsse, peter@maris.nl), who verifies if the institution is already registered. If a new entry is needed, the basic entry is made by MARIS, after which the appropriate NODC is responsible for updating further details and managing changes.

6.2. Data vocabulary

Use of common vocabularies in all meta-databases and data formats is an important prerequisite towards consistency and interoperability with existing Earth Observing systems and networks. Common vocabularies consist of lists of standardised Terms of Reference covering a broad spectrum of disciplines of relevance to the oceanographic and wider community. Using standardised ToR the problem of ambiguities related to data structure, organization and format is solved and therefore, common algorithms for data processing may be applied. This allows the interoperability of datasets in terms of their manipulation, distribution and long-term reuse.

ODYSSEA will adopt an Essential Variables list of terms (aggregated level) that has been defined and was published in June 2016 on the NERC/BODC Vocabulary Server[3].

This new vocabulary is mapped to the standards recommended for ODYSSEA parameter metadata: P01 (parameter), P07 (CF variable), P06 (units) from SeaDataNet controlled vocabularies managed by NERC/BODC and the internationally assured AphiaID from the WOrld Register of Marine Species (WoRMS)[4].

[3] https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/

[4] http://www.marinespecies.org/aphia.php?p=webservice

6.3. Metadata

Metadata refers to the description of datasets and services in a compliant form as it has been defined by the Directive 2007/2/EC (INSPIRE) and Commission Regulation No 1205/2008.

Metadata is the **data about the data**. Metadata describes how, when and by whom a particular set of data or a service was collected or prepared, and how the data is formatted, or the service is available. Metadata is essential for understanding the information stored in and has become increasingly important.

Metadata is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. Metadata is often called as the "data about the data or information about information".

Metadata is also data about services. Metadata describes the content, quality, condition, and other characteristics of a data set or the capabilities of a service. Creating metadata or data documentation for geospatial datasets is crucial to the data development process. Metadata is a valuable part of a dataset and can be used to:

- **Organize** data holdings (Do you know what you have?).
- Provide **information about** data holdings (Can you describe to someone else what you have?).
- Provide information **to data users** (Can they figure out if your data are useful to them?).
- **Maintain the value** of your data (Can they figure out if your data are useful 20 years from now?).

In the geographical domain we can have a description of spatial data (**spatial data** metadata), a service (**service** metadata) or a special analysis process (**process** metadata). Most for the standardization work is done for data metadata, however service and process metadata become increasingly important. Metadata is used in discovery mechanisms to bring spatial information providers and users together.

The following mechanisms are recognized:

- **Discovery**: which data source contains the information that I am looking for?
- **Exploration (or evaluation)**: do I find within the data sources the right information to suit my information needs?
- **Exploitation (use and access)**: how can I obtain and use the data sources?

Each mechanism has its own use of metadata. The selected standards should fulfil the needs to carry out services using these mechanisms. Metadata is required to provide information about an organisation's data holdings. Data resources are a major national asset, and information of what datasets exist within different organisations, particularly in the public sector, is required to improve efficiencies and reduce data duplication. Data catalogues and data discovery services enable potential users to find, evaluate and use that data, thereby increasing its value. This is also becoming important at the European level. In addition, metadata received from an external source may require further information supplied to metadata to allow easy process and interpretation.

In this context for all types of data the following information is required (SeaDataNet, 2010):

- **Where** the data were collected: location (preferably as latitude and longitude) and depth/height;
- **When** the data were collected (date and time in UTC or clearly specified local time zone);
- **How** the data were collected (e.g., sampling methods, instrument types, analytical techniques). How do we organize the data (e.g., in terms of station numbers, cast numbers);
- **Who** collected the data, including name and institution of the data originator(s) and the principal investigator;
- **What** has been done to the data (e.g., details of processing and calibrations applied, algorithms used to compute derived parameters);
- **Watch** points for other users of the data (e.g., problems encountered and comments on data quality).

The ICES Working Group on Data and Information Management (WGDIM) has developed a number of data type guidelines which itemize these elements that are required for thirteen different data types (see table below). These Data Type Guidelines have been developed using the expertise of the oceanographic data centres of ICES Member Countries. They have been designed to describe the elements of data and metadata considered as important to the ocean research community. These guidelines are targeted towards most physical-chemical-biological data types collected on oceanographic research vessel cruises. Each guideline addresses the data and metadata requirements of a specific data type.

This covers three main areas:

- What the data collector should provide to the data centre (e.g., collection information, processing, etc.);
- How the data centre handles data supplied (e.g., value added, quality control, etc.);
- What the data centre can provide in terms of data, referral services and expertise back to the data collector. A selection of these guidelines, in particular for those data types that are not yet dealt with in detail here, are included in Appendix 1 of this document.

This document summarizes the concept of metadata that is intended to be adopted by ODYSSEA data platform, following the commonly agreed INSPIRE data specification template in its relevant parts, i.e., dataset-level, services metadata and data quality. It also contains detailed technical documentation on the XML source-code level and therefore provides specific guidelines to correctly create and maintain metadata in the XML format.

6.4. Metadata Catalogue Service

A **Metadata Catalogue Service** is a mechanism for storing and accessing descriptive metadata and allows users to query for data items based on desired attributes. The catalogue service stores descriptive information (metadata) about logical data items. The Open Geospatial Consortium (OGC) has created the **Catalogue Service for Web (CSW) standard** to enable the easy data discovery from a catalogue node. Catalogue services support the ability to publish and search metadata for data, services, and related information. Metadata in catalogues can be queried and presented for evaluation and further processing

by both humans and software. Catalogue services (and other resources such as bibliographic resources, datasets, etc.) are required to support the discovery and binding to published web map services. The CSW standard is extremely rich. In addition to supporting a query from a user, it can support distributed queries (one query that searches many catalogues) and the harvesting of metadata from node to node.

Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects. Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community.

The International Organisation for Standardisation (ISO) includes ISO/TC 2112, which is an international, technical Committee for the standardisation of geographical information. TC 211 has created a strong, globally implemented set of standards for geospatial metadata: the baseline ISO 19115; ISO 19139 for implementation of data metadata and the ISO 19119 for services metadata.

These open standards define the structure and content of metadata records and are essential for any catalogue implementation. ISO 19115 describes all aspects of geospatial metadata and provides a comprehensive set of metadata elements. It is designed for electronic metadata services, and the elements are designed to be searchable wherever possible. It is widely used as the basis for geospatial metadata services. However, because of the large number of metadata elements and the complexity of their data model, implementation of ISO 19115 is difficult.

The INSPIRE DIRECTIVE applies these standards and specifications in its implementation. INSPIRE makes use of three catalogues for unique IDs management: **(1) SeaDataNet, (2) ICES and (3) CMEMS.** ICES catalogue has a geospatial component not present in the SeaDataNet catalogue while CMEMS provides the reference to model results.

6.4.1. Catalogue Service for Web (CSW)

This section describes briefly the Open GIS Consortium (OGC) specification for catalogue services. According to this specification: "*Catalogue services support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects; Metadata in catalogues represent resource characteristics that can be queried and presented for evaluation and further processing by both humans and software. Catalogue services are required to support the discovery and binding to registered information resources within an information community*".

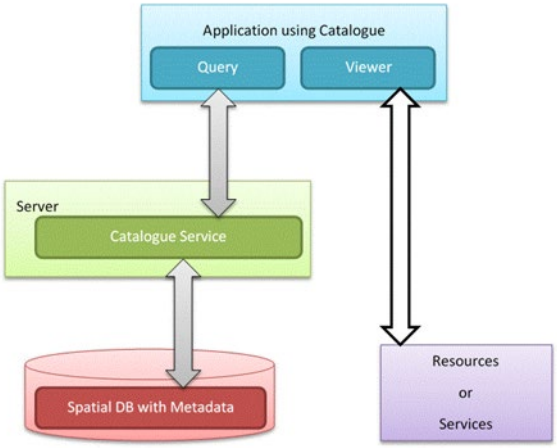


FIGURE 6.1: GENERIC VIEW OF THE CSW PROTOCOL AND ARCHITECTURE

The Inspire initiative uses the CSW protocol and the ISO metadata application profile (AP) for the specification and implementation of the Inspire Discovery Service. In ODYSSEA, the ODYSSEA ISO metadata profile will be developed and used as described in the metadata sections of this document.

6.4.2. Harvesting

Harvesting is the procedure of collecting metadata records from other (external) catalogues and synchronize the local catalogue with the collected information.

In the majority of the cases the harvesting process is scheduled and automatically executed once or at pre-defined intervals. It is usually also possible to execute a harvesting procedure on-demand, i.e., executed by human request.

The diagram below depicts a sample on how the harvesting procedures could be seen between the ODYSSEA platform catalogue and other external catalogues. To be noted that the harvesting procedure uses, within Inspire, the CSW protocol. Within the catalogue responses to the harvesting requests there are collections of metadata records, using the model described in this document (i.e., INSPIRE Datasets and Services).

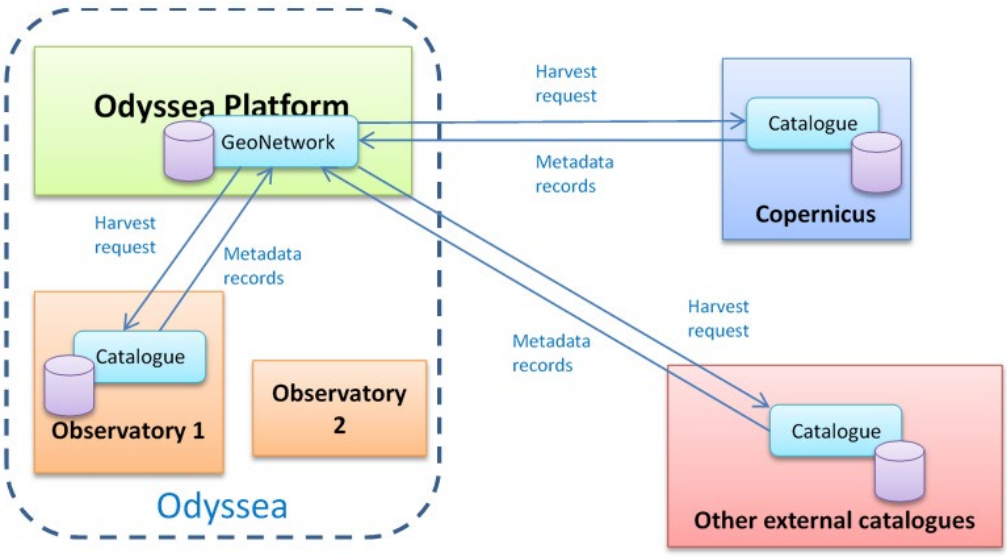


FIGURE 6.2: SAMPLE HARVESTING PROCEDURES BETWEEN ODYSSEA PLATFORM CATALOGUE AND EXTERNAL CATALOGUES.

6.5. Guidelines on using metadata elements

6.5.1. Lineage

Following the ISO 19113 Quality principles, if a data provider has a procedure for quality validation of their spatial datasets then the data quality elements, listed in Chapter 2, should be used. If not, the Lineage metadata element (defined in Regulation 1205/2008/EC) should be used to describe the overall quality of a spatial dataset.

According to Regulation 1205/2008/EC, lineage "is a statement on process history and/or overall quality of the spatial dataset. Where appropriate it may include a statement whether the dataset has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity. The value domain of this metadata element is free text".

Apart from describing the process history, if feasible within a free text, the overall quality of the dataset (series) should be included in the Lineage metadata element. This statement should contain any quality information required for interoperability and/or valuable for use and evaluation of the dataset (series).

6.5.2. Temporal reference

According to Regulation 1205/2008/EC, at least one of the following temporal reference metadata elements shall be provided: temporal extent, date of publication, date of last revision, date of creation. If feasible, the date of the latest revision of a spatial dataset should be reported using the date of latest revision in a metadata element.

6.5.3. Topic category

The topic categories defined in Part D.2 of the INSPIRE Implementing Rules for metadata are derived directly from the topic categories defined in B.5.27 of ISO 19115. Regulation 1205/2008/EC defines the INSPIRE data themes to which each topic category is applicable, i.e., oceanography is the INSPIRE theme for which the Geoscientific information topic category is applicable.

6.5.4. Keyword

Regulation 1205/2008/EC requires that, for a spatial dataset or a spatial dataset series, “at least one keyword shall be provided from the General Environmental Multi-lingual Thesaurus (GEMET) describing the relevant spatial data theme, as defined in Annex I, II or III to Directive 2007/2/EC”. Keywords should be taken from the GEMET – General Multilingual Environmental Thesaurus where possible.

7. ODYSSEA datasets

This section describes the structure and the content of the proposed ODYSSEA metadata profile on the dataset-level and includes general guidelines for the metadata from two points of view – the first one is the ODYSSEA metadata, while the second represents ODYSSEA data quality issues.

The structure described in this document is compliant with the existing ISO standards for metadata – i.e., especially ISO EN 19115 and ISO 19139. The full list of used ISO standards can be found in the List of References at the end of this document. The primary goal of this part of the deliverable is to develop a metadata profile for ODYSSEA geographic datasets and time-series datasets, within the framework of these ISO standards, aiding the support of the interoperability between the different metadata and/or GIS platforms.

The metadata model to be adopted in ODYSSEA is described in more detail in Annex I.

7.1. Dataset-level metadata

Metadata can be reported for each individual spatial object (spatial object-level metadata) or once for a complete dataset or dataset series (dataset-level metadata). If data quality elements are used at spatial object level, the documentation shall refer to the appropriate definition in the Data Quality Info section of this document. This section only specifies the dataset-level metadata elements.

For some dataset-level metadata elements, in particular on data quality and maintenance, a more specific scope can be specified. This allows the definition of metadata at sub-dataset level, e.g., separately for each spatial object type. When using ISO 19115/19139 to encode the metadata, the following rules should be followed:

- The scope element (of type DQ_Scope) of the DQ_DataQuality subtype should be used to encode the scope.
- Only the following values should be used for the level element of DQ_Scope: series, dataset, featureType.
- If the level is featureType[5] then the levelDescription/MD_ScopeDescription/features element (of type Set <GF_FeatureType>) shall be used to list the feature type names.
- Mandatory or conditional metadata elements are specified in the next sub-section, while optional metadata elements are specified in subsequent sub-Section. The tables describing the metadata elements contain the following information:
- The first column provides a reference to a more detailed description. The second column specifies the name of the metadata element.

[5] *The value featureType is used to denote spatial object types*

- The third column specifies the multiplicity.
- The fourth column specifies the condition, under which the given element becomes mandatory (only for the first and second tables).

In **Annex I** a detailed description of the metadata is presented.

7.2. Service-level metadata

This section describes the structure and the content of the proposed ODYSSEA metadata profile on the service-level and includes general guidelines for ODYSSEA metadata from two points of view – the first one is the ODYSSEA-specific metadata, while the second represents quality issues of the data published by the services.

The structure described in this document is compliant with the existing ISO standards for metadata – i.e., especially ISO EN 19115, EN ISO 19119 and ISO 19139 (the full list of used ISO standards can be found in List of References at the end of this document). The primary goal of this section is to explain the development in the metadata profile of ODYSSEA geographical data services, within the framework of these ISO standards. Through this process, the principle of interoperability is supported and data are easily harvested and exchanged between various discovery services and different metadata and/or GIS platforms.

Metadata can be reported for each individual spatial object (spatial object-level metadata) or once for a complete dataset or dataset series (dataset-level metadata). On the other hand, metadata can also be reported for the services that are publishing ODYSSEA data – i.e., especially INSPIRE view and download services. This section only specifies service-level metadata elements.

For some service-level metadata elements, in particular for data quality, a more specific scope can be specified. This allows the definition of metadata at sub-dataset level, e.g., separately for each spatial object type. When using ISO 19115/19139 to encode the metadata, the following rules should be followed:

- The scope element (of type DQ_Scope) of the DQ_DataQuality subtype should be used to encode the scope.
- Only the following value should be used for the level element of DQ_Scope: service.

Mandatory or conditional metadata elements are specified in the ANNEX I. Optional metadata elements are specified in the subsequent sub-section of this ANNEX.

7.3. Data format standards

7.3.1. Ocean Data View data model and netCDF Format

As part of the ODYSSEA services, data sets will be accessible via download services. Delivery of data to users requires common data transfer formats, which interact with other standards (Vocabularies, data quality control). In SeaDataNet it was decided that Ocean Data View (ODV) and netCDF format are mandatory.

ODYSSEA will follow the SeaDataNet (2017) procedures, as main concepts of this document are reproduced in the following paragraphs. ODYSSEA will also follow the fundamental data model underlying ODV format which, in practice, is composed of a collection of rows, each having the same fixed number of columns.

In this model there are three different types of columns:

- The metadata columns;
- The primary variable data columns (one column for the value plus one for the qualifying flag);
- The data columns.

The metadata columns are stored at the left-hand end of each row, followed by the primary variable columns and then the data columns.

There are three different types of rows:

- The comment rows;
- The column header rows;
- The data rows.

The CF metadata conventions (http://cf-pcmdi.llnl.gov/) are designed to promote the processing and sharing of data files created with the NetCDF API. The conventions define metadata that provide a definitive description of what the data in each variable represents, and the spatial and temporal properties of the data. This enables users of data from different sources to decide which quantities are comparable, and facilitates building applications with powerful extraction, re-gridding, and display capabilities.

The standard is both mature and well-supported by formal governance for its further development. The standard is fully documented by a PDF manual accessible from a link from the CF metadata homepage (http://cf-pcmdi.llnl.gov/). Note that CF is a developing standard and consequently access via the homepage rather than through a direct URL to the document is recommended to ensure that the latest version is obtained. The current version of this document was prepared using version 1.6 of the conventions dated 5 December 2011.

The approach taken with the development of the SeaDataNet profile based on CF 1.6 was to classify data on the basis of feature types and produce a SeaDataNet specification for storage of each of the following:

- **Point time series**, such as current meter or sea level data, have row_groups made up of measurements from a given instrument at different times. The metadata date and time are set to the time when the first measurement was made. The primary variable is time (UT) encoded either as:
    - A real number representing the Chronological Julian Date, which is defined as the time elapsed in days from 00:00 on January 1st 4713 BC. If this option is chosen, then the column must have the heading 'Chronological Julian Date [days]'.
    - A string containing the UT date and time to sub-second precision corresponding to ISO8601 syntax (YYYY-MM-DDThh:mm:ss.sss) for example 2009-02-12T11:21:10.325. If

this option is chosen, the column must have the heading 'time_ISO8601'. If the time is not known to sub-second precision, then use the ISO8601 form appropriate to the known precision. For example, a timestamp to the precision of one hour would be represented by 2009-02-12T11:00 and a time stamp to a precision of a day by 2009-02-12.

Rows within the row_group are ordered by increasing time. Note that the z co-ordinate (e.g., instrument depth), essential for many types of time series data, needs to be stored as a data variable and could have the same value throughout the row_group.

- **Profile data**, such as CTD or bottle data, have row_groups made up of measurements at different depths. The metadata date and time are set to the time when the profile measurement started. The primary variable is the 'z co-ordinate', which for SeaDataNet is either depth in metres or pressure in decibars. Rows within the row_group are ordered by increasing depth.
- **Trajectories**, such as underway data, have row_groups made up of a single measurement, making the metadata time and positions the spatio-temporal co-ordinate channels. The primary variable is the 'z co-ordinate', which for SeaDataNet is standardised as depth in metres. Rows within the row_group are ordered by increasing time;
- **TimeSeriesProfile** (x, y, z fixed; t variable) but some variables can be measured at different depths at the same time var=f(t, z). The specification given is for storage of time series profiles such as moored ADCP.
- **TrajectoryProfile** (x, y, z, t all variable) but some variables can be measured at different depths at the same time var=f(t, z). The specification given is for storage of trajectory profiles such as shipborne ADCP.

The specification was then developed through discussions on a collaborative e-mail list involving participants in SeaDataNet, MyOcean, USNODC, NCAR and AODN. The working objective focussed on producing profiles with the following properties:

- CF 1.6 conformant;
- Have maximum interoperability with CF 1.6 implementations in use by MyOcean (OceanSITES conventions), USNODC (USNODC NetCDF templates) and two contributors to AODN (IMOS and METOC);
- Include storage for all labels, metadata and standardised semantic mark-up that were included in the SeaDataNet ODV format files for the equivalent feature type.

Significant list discussion focussed on the version of netCDF that should be used for SeaDataNet. The conclusion was that netCDF 4 should be used wherever possible, but that netCDF 3, although strongly discouraged, should not be totally forbidden.

On ANNEX II some examples of the structure of these files are presented.

7.3.2. Static data (Bathymetric, Chemical, Geologic, Geophysical, Biological, Biodiversity data)

ODYSSEA will also adopt the SeaDataNet proposed standards for marine chemistry (to support the EMODNet Chemistry pilot), bathymetry (to support the EMODNet Hydrography and Seabed Mapping

pilots), and geology and geophysics (to support the Geo-Seas project and the EMODNet Geology pilot). and marine biology.

Based on an analysis of the present situation, and currently existing biological data standards and initiatives, such as the Ocean Biogeographic Information System (OBIS), Global Biodiversity Information Facility (GBIF), Working Group on Biodiversity Standards (TDWG) and World Register of Marine Species (WoRMS) standards, SeaDataNet proposed a format for data exchange of biological data.

Key issues that steered the format development were (SeaDataNet III, publishable summary):

- Requirements posed by the intended use and application of the data format (data flows, density calculations, biodiversity index calculations, community analysis, etc…)
- Availability of suitable vocabularies (World Register of Marine Species, SeaDataNet Parameter list, SeaDataNet Unit list, etc…)
- Requirements for compatibility with existing tools and software (WoRMS taxon match services, EurOBIS QC services, Lifewatch workflows, Ocean Data View, etc…)
- The requirements of the extended ODV format for biological data were defined as follows:
- The format should be a general and higher level format without necessarily containing all specifics of each data type, but rather focusing on common information elements for marine biological data.
- At the same time the format needs to be sufficiently flexible/extendable to be applicable for at least part of the variety of biological data the NODC's are managing.
- It should be possible to derive OBIS or Darwin Core compatible datasets from the format.
- The format should be self-describing, in the sense that all information needed to interpret the data should be included in the file format or be available through links to vocabularies or term lists that are part of the format.

A specific ODV extended format for biological data has been defined for different types of files such as (see for details SeaDataNet deliverable D8.4):

- macrobenthos community with density and biomass values;
- zooplankton community with samples from different depths;
- demersal fish population with densities for different size classes and individual fish measurements;
- pollutant concentrations in biota specimens.

7.3.3. Open source Semantic Information

Semantic information may be useful for a myriad of services to the end users. However, the sources providing semantically rich information are very heterogeneous. Semantically rich information can be found on Wikipedia and Wikidata for instance. EMODnet, through the "Human activities" data sets, also provides some semantically rich information.

As one can see the sources of semantically rich information are very heterogeneous in their availability, reliability and format. Furthermore, they provide heterogeneous and partially redundant information. No

standard model exists for that type of information, as their variability is very high. However, as one of ODYSSEA platform aim is to integrate and fuse this kind of information, one must rely on a shared format in order to analyze and make use of it.

Within the services that will be developed in ODYSSEA, a domain ontology will be used in order to enable the integration of semantic information sources. **For each ODYSSEA use case, and for each ODYSSEA product relying on semantic information analysis and integration, end users of the products will have to develop, together with ODYSSEA technical partners, an ontology defining the concepts of interest of the use case.** This ontology will be the pivot language and representation format used to integrate heterogeneous open information sources.

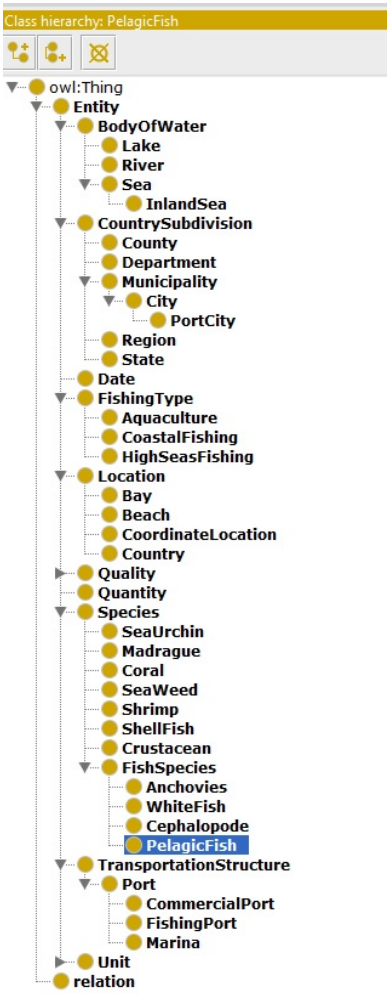


FIGURE 7.1: EXAMPLE OF AN ONTOLOGY DEFINING THE MAIN CONCEPTS USED TO ANALYZE THE IMPACT OF PORT STRUCTURES ON THE QUALITY OF BATHING WATERS AND FISH PRODUCTION

8. Data privacy policy

8.1. General principles

Basic principles regulated by the Data Protection Act will be observed namely:

- ODYSSEA will only hold the personal data which is necessary to offer services provided by its platform.
- Data is only used for the purposes described in the Data Protection Register Form and the Informed Consent Form.
- Personal data will only be held for as long as necessary. Once data are no longer needed it will be deleted from ODYSSEA records by the ODYSSEA platform Administrator (namely the CLS Chief Technical Officer (CTO) / IT platform manager). More specifically, in case a certain period (one year) is passed without the entry of an end-user in the platform, CLS will alert him through a standardized electronic message on the destruction of personal data.
- Personal data storage will be secured to ensure that data are not accessible to unwanted third parties and are protected against disaster and risk.
- ODYSSEA will regularly email website news and information updates only to those end-users and customers who have specifically subscribed to our email service. All subscription emails sent by the ODYSSEA platform will contain clear information on how to unsubscribe from our email service.
- In any event, no personal data will be shared with any third party for direct marketing. ODYSSEA will never sell, rent or exchange mailing lists of personal data.
- All ODYSSEA partners shall comply with the data protection and privacy laws applicable in their country of origin, including their national laws applicable to exporting data into the EU.
- ODYSSEA partners from non-EU countries have provided signed declarations that they will meet all relevant H2020 ethical standards and regulations. Exporting personal data from the EU to non-EU countries must comply with the applicable EU rules on cross-border transfer of personal data.
- In accordance with the Privacy and Electronic Communications (EC Directive) Regulations 2003, ODYSSEA will never send bulk unsolicited emails, (popularly known as Spam) to any email addresses.
- ODYSSEA may send emails to existing end-users and customers or prospective end-users and customers having inquired or registered in the ODYSSEA platform, regarding products or services directly provided by the ODYSSEA platform.
- All emails sent by ODYSSEA will be clearly marked as originating from this platform. All such emails will also include clear instructions on how to unsubscribe from ODYSSEA email services. Such instructions will either include a link to a page to unsubscribe or a valid email address to which the user should reply, with “unsubscribe” as the email subject heading.

Details on the protection of end-users’ personal data and the privacy rules to be followed by ODYSSEA, the participation of non-EU countries and the Informed Consensus Procedures are provided in Deliverable 1.1.

8.2. Use of Cookies

Cookies are small text files which are placed on your computer by websites that you visit. They are widely used in order to make websites work, or work more efficiently, as well as to provide information to the owner of the site.

ODYSSEA platform may generate cookies in order to work more efficiently. These will enhance features such as platform search and optimized page loading.

ODYSSEA may use Google Analytics to collect quantitative information on platform's performance and end-user's interaction with the platform. ODYSSEA will use this information to improve the service and experience offered by the platform. The use of Social Media buttons on some of the pages link to third party websites and services, like Facebook and Twitter also create cookies. These services use cookies when clicking the button. Privacy policies will be available for all these services and users should be able to read them to be informed on how their information is being used, and how they can opt-out, should they wish to.

9. References

AtlantOS, 2016, *Data Management Handbook*

Carval Thierry, 2016, *Catalogue of data and platforms at Network GDAC level, including the example of Copernicus In Situ TAC*. IMN/IDM/ISI/FM/16-024. http://doi.org/10.13155/45063

Columbus Project, 2017, *Use and sharing of marine observations and data by industry*

Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

Corrigendum to Commission Regulation (EC) No 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata

Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE)

Draft Commission Regulation (EU) NoL/... of [L] Implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial s and services

EN ISO 19113:2002 Geographic information – Quality principles

EN ISO 19114:2003/Corr. 1 :2005 Geographic information – Quality evaluation procedures

EN ISO 19115:2003/19115-1 :2014: Geographic information – Metadata

EN ISO 19115:2003/Cor. 1:2006

EN ISO 19108: Geographic information – Temporal schema

EN ISO 19108:2002/Cor. 1:2006

EN ISO 19119:2005/19119 :2016: Geographic information - Services

EN ISO 8601:2004 Data elements and interchange formats – Information exchange – Representation of dates and times

INSPIRE DS-D2.3, Definition of Annex Themes and Scope, v3.0,

INSPIRE DS-D2.5, Generic Conceptual Model, v3.4,

INSPIRE DS-D2.6, Methodology for the development of data specifications, v3.0,

INSPIRE DS-D2.7, Guidelines for the encoding of spatial data, v3.1,

INSPIRE DS-D2.8.I.7, INSPIRE Data Specifications on Transport Networks – Guidelines, v3.0

INSPIRE DS-D2.8.I.8, INSPIRE Data Specifications on Hydrography – Guidelines, v3.0

INSPIRE DS-D2.8.I.9, INSPIRE Data Specifications on Protected Sites – Guidelines, v3.0

INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, v. 1.1

ISO/TS 19138:2006 Geographic information – Data quality measures

ISO/TS 191139:2007 Geographic information – Metadata – XML Schema Implementation

SeaDataNet, 2010, *Data Quality Control Procedures*

SeaDataNet, 2017, *Data File Formats: ODV, Medatlas, netCDF. Deliverable D8.5*

ANNEX I. Data Ingestion service manual

The data ingestion component is responsible for acquiring data from local observatories and other external data sources into the ODYSSEA platform. Data is downloaded and converted into a standard format (netCDF following SeaDataNet guidelines). These standardized files are made available in a network share so that other parts of the ODYSSEA platform can use them (SOS, WMS, etc). This service manual explains how to setup and maintain the data ingestion components.

A-I.1. Data Ingestion component docker images

A-I.1.1.Functional requirements

The data ingestion component is made up of four separate docker images:

- Download Worker
    - Performs the download according to the configuration (remote address, data type, etc) that resides in the *Configuration Storage*. The Download and conversion progress is saved back on to the *Configuration Storage*. The docker image contains the .net core Runtime 2.0, Java runtime (java-1.8.0-openjdk) and the Nemo 1.6.6 program (https://www.seadatanet.org/Software/NEMO). Since the download worker immediately calls the database on start-up and the docker compose does not guarantee each container is fully operational before starting the next one (https://docs.docker.com/compose/startup-order/), it tries to connect to the database several times on start-up.
- Configuration storage
    - This docker image contains a PostgreSQL database with the necessary schema for the application.
- Configuration API
    - The configuration API is a JSON web-service to perform CRUD operation on the download jobs, trigger and reschedule existing jobs and examine the logs of each of the performed downloads. The docker image contains the .net core runtime 2.0, apache server 2.4.33 and supervisor (http://supervisord.org/) to start both apache and the application. Apache is needed as a proxy (https://docs.microsoft.com/en-us/aspnet/core/host-and-deploy/linux-apache?view=aspnetcore-2.1&tabs=aspnetcore2x).
- Configuration UI
    - The Configuration UI is an Angular web application that consumes the web-services exposed by the Configuration API.

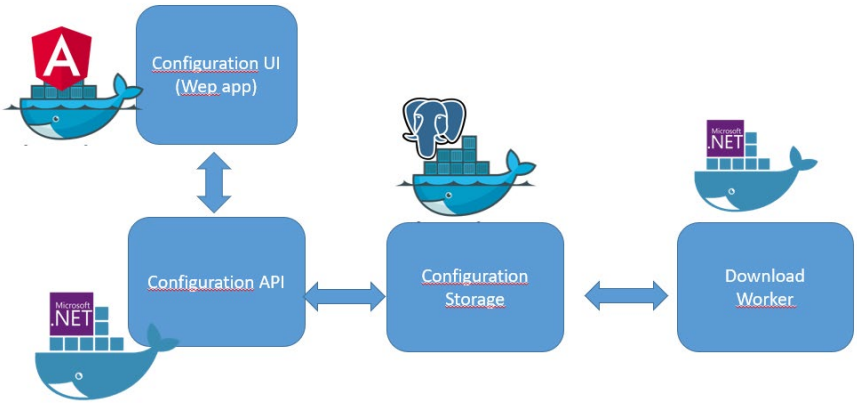


FIGURE A-I.1: THE ODYSSEA DATA INGESTION COMPONENT.

All docker images are based on the CLS Centos 7 image:

registry-ext.cls.fr:443/docker-cls/os/centos7:clslatest

The next sections explain each of the docker images in detail.

A-I.2. Configuration Storage docker image

The only software installed on the original docker image supplied by CLS was the PostgreSQL 9.4 database server. The necessary database schemas are then uploaded into the docker image. An intermedia docker image created just with the PostgreSQL server can be accessed at:

registry-ext.cls.fr:443/pgalvao/dataingestion/centos7-cls-postgres

The DockerFile that creates this image is available at:

https://gitshare.cls.fr/pgalvao/dataingestion/tree/master/CLS_DockerSetup/pgDatabase/

Note that this is a generic docker and misses the necessary databases for the Data ingestion component. We are mentioning it in this service manual just in case other partners need a PostgreSQL database server.

This intermediate docker image is used to create the Configuration Storage docker image. The docker file that performs this operation is available at:

https://gitshare.cls.fr/pgalvao/dataingestion/blob/master/OdysseaDownloadBackend/Dockerfile.dpl.ods.DbPsql

The final docker image that is used in the Data ingestion component is available at:

registry-ext.cls.fr:443/pgalvao/dataingestion/c7.psql.download.db

To integrate this component with the following environment variables it should be passed with docker create or compose:

- DB_USER: odyssea
- DB_PASS: odyssea123+
- DB_NAME: odss_dwd_settings
- DB_NAME2: odss_dwd_jobs

This will create the two databases that are used by the application: one with the settings and another used by the Hangfire automation server (see next section for more details).

A-I.3. Download Worker docker image

The *Download Worker* performs most of the heavy lifting in the application. It performs the download and the conversion of external data according to the definitions stored in the *odss_dwd_settings* database that is server by the Configuration storage component.

The application was implemented in .net core (https://docs.microsoft.com/en-us/dotnet/core/index) and uses Hangfire (https://www.hangfire.io/). It has a background job scheduler. The Hangfire library uses the *odss_dwd_jobs* database server by the Configuration storage component. Originally this database is empty since the library creates the schema on demand.

Once the data is downloaded and parsed the Nemo software (https://www.seadatanet.org/Software/NEMO) is used to create standard netCDF files. This is a java-based application.

The *Download Worker* docker image installs over the *CLS Centos 7*:

- DotnetCore Runtime 2.0
- java-1.8.0-openjdk
- Part of Nemo software (nemo is a UI based application, so some adaptations where made).

Some intermediate docker images where created namely:

- registry-ext.cls.fr:443/pgalvao/dataingestion/centos7-cls-netsdk
    - This image contains the full .netcore SDK. This is only necessary for compilation or software development. We created this component to use in the CI/CD system, so that the application is compiled in the save version of the operational system that will run it.

The full SDK is quite big about 1GB, but again this image is not used in the final component. The docker file to create this component is at: https://gitshare.cls.fr/pgalvao/dataingestion/blob/master/CLS_DockerSetup/dotNET/Dockerfile.sdk

- registry-ext.cls.fr:443/pgalvao/dataingestion/centos7-cls-netrt
    - Image with the .net runtime (about 300 Mb). The net runtime must be installed in any image that will run .net code. The docker file that generates this image is at: https://gitshare.cls.fr/pgalvao/dataingestion/blob/master/CLS_DockerSetup/dotNET/Dockerfile.rt
- registry-ext.cls.fr:443/pgalvao/dataingestion/centos7-cls-netrt-nemo
    - Image with .net runtime, Java runtime and the nemo executable. The docker file to create this image is at: https://gitshare.cls.fr/pgalvao/dataingestion/blob/master/CLS_DockerSetup/dotNET/Dockerfile.nemo.rt

This final image will be the one used for creating the *Download Worker* image.

---

registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.server

---

A-I.4. Configuration API image

The configuration API was implemented in .netCore. The base image with .net runtime described in the previous section is used. The API will perform the CRUD operation on both the *odss_dwd_settings* and *odss_dwd_jobs* database server by the Configuration storage.

The final image can be accessed at:

---

registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.config.api

---

In order to accept requests from outside the docker container, this image needs a webserver installed. In this case apache was used. The web server is used as a reverse proxy to forward the requests to the .net application (https://docs.microsoft.com/en-us/aspnet/core/host-and-deploy/linux-apache?view=aspnetcore-2.1&tabs=aspnetcore2x). Since each docker image should only use a single entry point (and in this case we need to start both the application and the apache server), the supervisor control system was installed (http://supervisord.org/) and used as an entry point that starts both the apache server and the configuration API.

A-I.5. Configuration UI

The configuration UI was built as a front-end to the services exposed by the *Configuration API*. Angular was used to develop the application, so once its compiled doesn't need any frameworks on the base image. Currently this is the only component that is not using the *CLS Centos 7* as a base image. It is based on the nginx:alpine image that is only a few megabytes in size (less than 30). This will be updated in the next iteration. The final version is available here:

registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.config.ui

A-I.5.1.Composing the application

The data ingestion component is composed by 4 docker images:

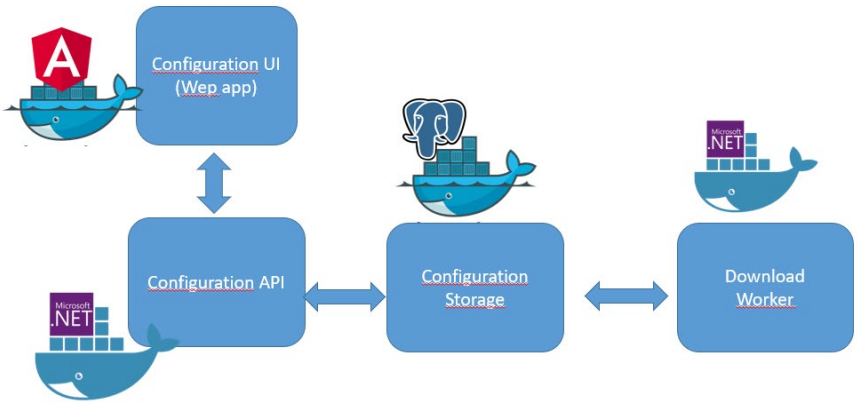


FIGURE A-I.2: THE FOUR DOCKERS OF THE ODYSSEA DATA INGESTION COMPONENT.

- Download Worker
  - docker pull registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.server
- Configuration storage
  - docker pull registry-ext.cls.fr:443/pgalvao/dataingestion/c7.psql.download.db
- Configuration API
  - docker pull registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.config.api
- Configuration UI
  - docker pull registry-ext.cls.fr:443/pgalvao/dataingestion/c7.download.config.ui

The docker-compose file that integrates the different docker is available at:

https://gitshare.cls.fr/pgalvao/dataingestion/tree/master/docker-compose.yml

The compose file that maps the *Configuration Store* has the following structure:

```
db:

    image: registry-
ext.cls.fr:443/pgalvao/dataingestion/c7.psql.download.db:latest

    environment:

      DB_USER: odyssea

      DB_PASS: odyssea123+

      DB_NAME: odss_dwd_settings

      DB_NAME2: odss_dwd_jobs

    ports:

      - "5440:5432"
```

The database options (user credentials and database names) should not be modified.

The *Configuration API* that is mapped has the following structure:

```
odysseadownloadwebapi:

    image: registry-
ext.cls.fr:443/pgalvao/dataingestion/c7.download.config.api:latest

    ports:

      - "338:80"

    depends_on:

        - db
```

This component needs to map a port to the server so that the API is accessible.

The *Download Worker* that is mapped has the following structure:

```
odysseadownloadserver:

    image: registry-
ext.cls.fr:443/pgalvao/dataingestion/c7.download.server:latest

    volumes:
```

```
    - type: volume
      source: downloaded
      target: /downloaded
  depends_on:
      - db
```

Please make sure to update the file with the correct target instruction for the host machine, where the application is deployed. This is where the generated netCDF files are stored.

A-I.5.2.Running the docker compose

After running the docker compose on the file available on https://gitshare.cls.fr/odyssea/data-collection/blob/master/docker-compose.yml three docker containers will be set up and running:

1. A container with the configuration database

   registry-ext.cls.fr:443/odyssea/data-collection/c7.psql.dataingestion.db

   1. A container with the download server

      registry-ext.cls.fr:443/odyssea/data-collection/c7.dataingestion.server

   2. A container with a configuration API

      registry-ext.cls.fr:443/odyssea/data-collection/c7.dataingestion.api

To test the gloss download, navigate to the address of the configuration API container (in my case http://localhost:337). A SWAGGER UI is available to test the available API Calls.

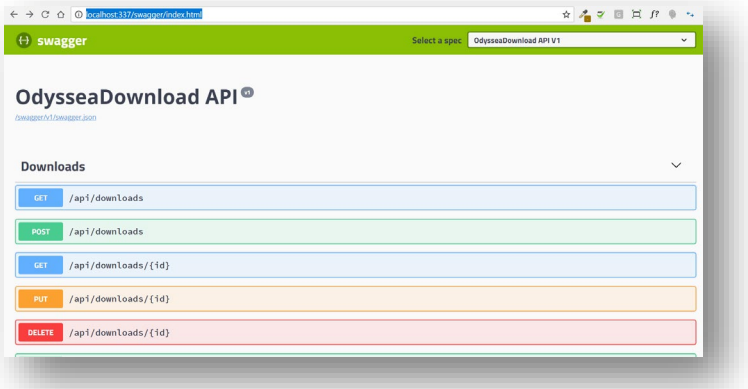
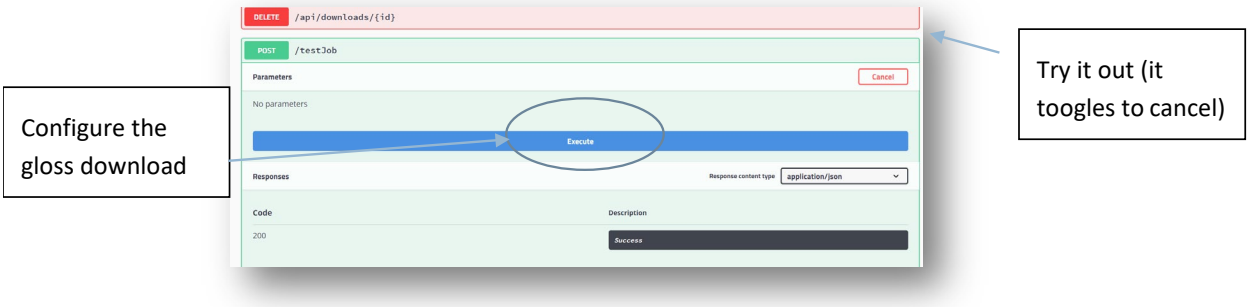
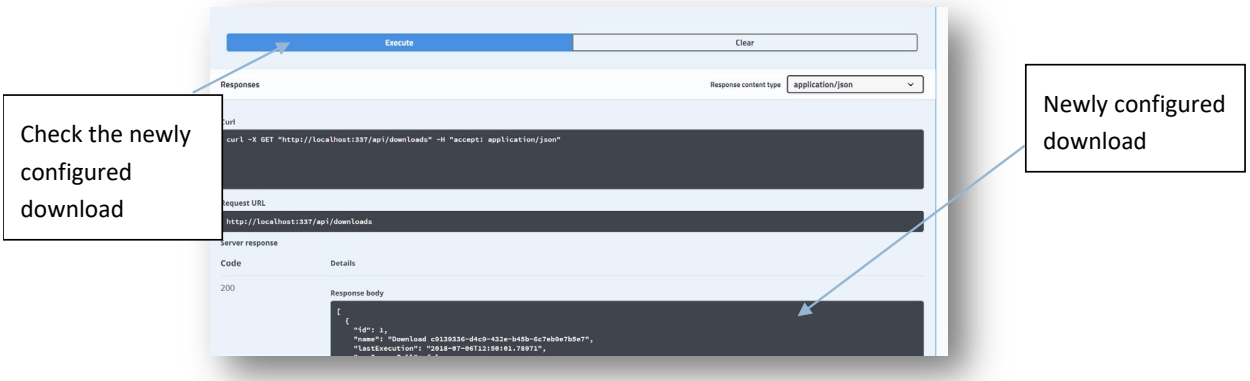


FIGURE A-I.3: SWAGGER USER INTERFACE

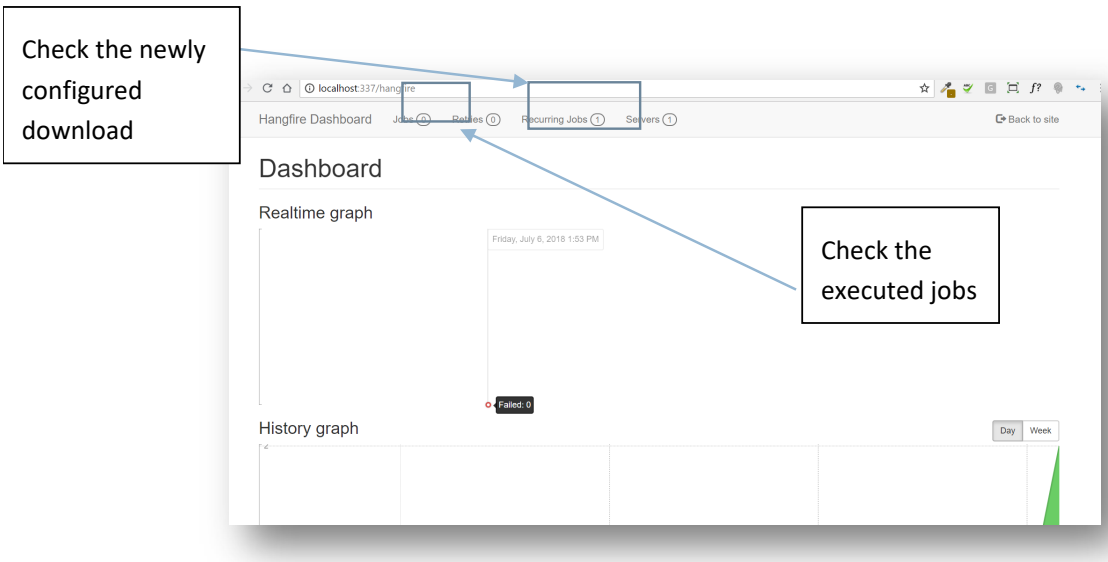
Expand the last method and execute it, select try it out and then execute. It should return a status code of 200.



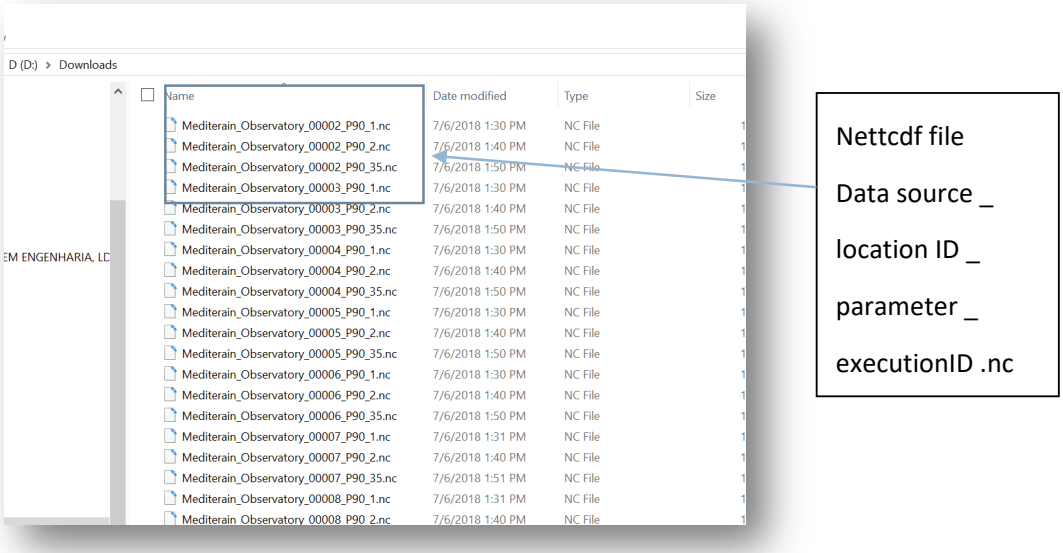
Running the same procedure on the api/download call will then return the newly configured download.



To monitor the download progress, go to the configuration API address / Hangfire (in my case http://localhost:337/hangfire)



As the jobs are executed, the netCDF files are being generated on the mapped folder of the odysseadownloadserver in the docker file. In my case it's the d:\downloads



ANNEX II. Mandatory and optional metadata elements

The metadata describing a spatial dataset or a spatial dataset series shall comprise the metadata elements required by Commission Regulation No 1205/2008/EC (implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata) for spatial datasets and spatial dataset series (*cf. Table A-II.2*) as well as the ODYSSEA-specific mandatory metadata elements specified in *Table A-II.3* and optionally elements specified in *Table A-II.4*.

A-II.1. Mandatory metadata elements

TABLE A-II.2: METADATA FOR SPATIAL DATASETS AND SPATIAL DATASET SERIES SPECIFIED IN THE INSPIRE METADATA REGULATION [REGULATION 1205/2008/EC].

Metadata			
Regulation Section	**Metadata element**	**Multiplicity**	**Condition**
1.1	Resource title	1	
1.2	Resource abstract	1	
1.3	Resource type	1	
1.4	Resource locator	0..*	Mandatory if a URL is available to obtain more information on the resource, and/or access related services.
1.5	Unique resource identifier	1..*	
1.7	Resource language	0..*	Mandatory if the resource includes textual information.
2.1	Topic category	1..*	
3	Keyword	1..*	
4.1	Geographic bounding box	1..*	
5	Temporal reference	1..*	
6.1	Lineage	1	
6.2	Spatial resolution	0..*	Mandatory for datasets and dataset series if an equivalent scale or aresolution distance can be specified.
7	Conformity	1..*	
8.1	Conditions for access and use	1..*	
8.2	Limitations on public access	1..*	
9	Responsible organisation	1..*	
10.1	Metadata point of contact	1..*	
10.2	Metadata date	1	
10.3	Metadata language	1	

TABLE A-II.3: MANDATORY AND CONDITIONAL THEME-SPECIFIC METADATA FOR THE ODYSSEA METADATA PROFILE

ODYSSEA metadata profile Section	Metadata element	Multiplicity	Condition
1	Coordinate reference system	1	
2	Encoding	1..*	
3	Character Encoding	0..*	Mandatory, if a non-XML-based encoding is used that does not support UTF-8
4	Source title	0..1	Mandatory, if a source title can be specified
5	Source date of mapping	0..*	Mandatory, if source date of mapping can be specified

A-II.1.1.Coordinate Reference System

Metadata element name	Coordinate Reference System
Definition	Description of the horizontal coordinate reference system used in the dataset.
ISO 19115 number and name	13. referenceSystemInfo
ISO/TS 19139 path	referenceSystemInfo
INSPIRE obligation / condition	mandatory
INSPIRE multiplicity	1
Data type (and ISO 19115 no.)	189. MD_CRS
Domain	Either the referenceSystemIdentifier (RS_Identifier) or the projection (RS_Identifier), ellipsoid (RS_Identifier) and datum (RS_Identifier) properties shall be provided.
Implementing instructions	–
Example	referenceSystemIdentifier: code: ETRS_89 codeSpace: INSPIRE RS registry
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:referenceSystemInfo> <gmd:MD_ReferenceSystem> <gmd:referenceSystemIdentifier> <gmd:RS_Identifier> .. <gmd:code> <gco:CharacterString>**3035**</gco:CharacterString> </gmd:code> <gmd:codeSpace> <gco:CharacterString>**EPSG**</gco:CharacterStr ing> </gmd:codeSpace> ..

	</gmd:RS_Identifier> </gmd:referenceSystemIdentifier> </gmd:MD_ReferenceSystem> </gmd:referenceSystemInfo> .. </gmd:MD_Metadata>
Comment	To ensure unified approach of the codes across the whole Europe, we recommend an approach that uses standardize EPSG codes for the expression of the reference system (like 4326 stands for the coordinate system WGS84 or 3035 for ETRS89).

A-II.1.2.Encoding

Metadata element name	Encoding
Definition	Description of the computer language construct that specifies the representation of data objects in a record, file, message, storage device or transmission channel.
ISO 19115 number and name	271. distributionFormat
ISO/TS 19139 path	distributionInfo/MD_Distribution/distributionFormat
INSPIRE obligation / condition	mandatory
INSPIRE multiplicity	1..*
Data type (and ISO 19115 no.)	284. MD_Format
Domain	The following property values shall be used for default and alternative encodings specific and developed in the ODYSSEA project: Default Encoding name: ODYSSEA version: version 3.0; GML, version 3.2.1
Implementing instructions	–
Example	name: ODYSSEA version: version 3.0, GML, version 3.2.1
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:distributionInfo> <gmd:MD_Distribution> <gmd:distributionFormat> <gmd:MD_Format> <gmd:name> <gco:CharacterString>**GML**</gco:CharacterString> </gmd:name> <gmd:version> <gco:CharacterString>**3.2.1**</gco:CharacterString> </gmd:version> </gmd:MD_Format> .. </gmd:distributionFormat> </gmd:MD_Distribution> </gmd:distributionInfo> .. </gmd:MD_Metadata>
Comment	–

A-II.1.3.Character Encoding

Metadata element name	Character Encoding
Definition	Full name of the character coding standard used for the dataset
ISO 19115 number and name	4. characterSet
ISO/TS 19139 path	identificationInfo/*/characterSet
INSPIRE obligation / condition	Mandatory, if a non-XML-based encoding is used that does not support UTF-8
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	40. MD_CharacterSetCode
Domain	Codelist (See B.5.10 of ISO 19115)
Implementing instructions	–
Example	–
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:identificationInfo> .. <gmd:characterSet> <gmd:MD_CharacterSetCode codeList="**http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmxCodelists.xml#MD_ClassificationCode**" codeListValue="utf8">**utf8**</gmd:MD_CharacterSetCode> </gmd:characterSet> .. </gmd:identificationInfo> .. </gmd:MD_Metadata>
Comment	-

A-II.1.4.Source title

Metadata element name	Source title
Definition	Full name by which the input data source of the described dataset is known
ISO 19115 number and name	360. title
ISO/TS 19139 path	dataQualityInfo/*/lineage/*/source/*/sourceCitation/*/title
INSPIRE obligation / condition	0..1
INSPIRE multiplicity	Mandatory, if a source title can be specified.
Data type (and ISO 19115 no.)	96. sourceCitation
Domain	Free text

Implementing instructions	Full name of the source dataset without an expression of the scale – a scale (in a form of a denominator) should be expressed in section 4.2.4.
Example	Map of pedogenetic associations, ODYSSEA Dataset
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:dataQualityInfo> .. <gmd:lineage> <gmd:LI_Lineage> <gmd:source> <gmd:LI_Source> .. <gmd:sourceCitation> <gmd:CI_Citation> <gmd:title> <gco:CharacterString>**ODYSSEA maritime dataset**</gco:CharacterString> </gmd:title> .. </gmd:CI_Citation> </gmd:sourceCitation> </gmd:LI_Source> </gmd:source> .. </gmd:LI_Lineage> </gmd:lineage> </gmd:dataQualityInfo> .. </gmd:MD_Metadata>
Comment	–

A-II.1.5.Source date of mapping

Metadata element name	Source date of mapping
Definition	Reference date for the input data source of the described dataset.
ISO 19115 number and name	362. date
ISO/TS 19139 path	dataQualityInfo/*/lineage/*/source/*/sourceCitation/*/date/*/date
INSPIRE obligation / condition	Mandatory, if a source date of mapping can be specified
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	393. CI_Date
Domain	Described in ISO 19108 and ISO 8601
Implementing instructions	This metadata should be filled with these elements: - date (e.g. 2010, 2010-04, 2010-04-09) - dateType (i.e. creation,revision or publication)
Example	–
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:dataQualityInfo> ..

	<gmd:lineage> <gmd:LI_Lineage> <gmd:source> <gmd:LI_Source> .. <gmd:sourceCitation> <gmd:CI_Citation> .. <gmd:date> <gmd:CI_Date> <gmd:date> <gco:Date>**2010-04-09**</gco:Date> </gmd:date> <gmd:dateType> <gmd:CI_DateTypeCode codeList="**http://www.isotc211.org/2005/resources/codeList.xml#CI_DateTypeCode**" codeListValue="**creation**" /> </gmd:dateType> </gmd:CI_Date> </gmd:date> .. </gmd:CI_Citation> </gmd:sourceCitation> </gmd:LI_Source> </gmd:source> </gmd:LI_Lineage> </gmd:lineage> </gmd:dataQualityInfo> .. </gmd:MD_Metadata>
Comment	–

A-II.2. Optional metadata elements

The metadata describing a spatial dataset or a spatial dataset series related to ODYSSEA should comprise the ODYSSEA-specific metadata elements, specified in Table 4. This Table contains all metadata elements that have a multiplicity of 0..1 or 0..*.

TABLE A-II.4: OPTIONAL THEME-SPECIFIC METADATA FOR THE ODYSSEA METADATA PROFILE

Metadata profile section	Metadata element	Multiplicity
1	Spatial representation type	0..*
2	Online digital transfer options	0..*
3	Topology level	0..1
4	Source mapping scale	0..1
5	Data Quality – Completeness - Omission	0..*
6	Data Quality – Positional accuracy – Absolute or external accuracy	0..*
7	Data Quality – Thematic accuracy – Classification correctness	0..*

A-II.2.1.Spatial representation type

Metadata element name	Spatial representation type
Definition	Method used to spatially represent geographic information.
ISO 19115 number and name	37. spatialRepresentationType
ISO/TS 19139 path	identificationInfo/*/spatialRepresentationType
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	MD_SpatialRepresentationTypeCode
Domain	Codelist (See B.5.26 of ISO 19115)
Implementing instructions	–
Example	–
Example XML encoding	<gmd:MD_Metadata .. .. <gmd:identificationInfo> .. **<gmd:spatialRepresentationType>** **<gmd:MD_SpatialRepresentationTypeCode** **codeList**="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmx Codelists.xml#MD_SpatialRepresentationType Code" codeListValue="vector">vector</gmd:MD_SpatialRepresentationTypeCode> </gmd:spatialRepresentationType> .. </gmd:identificationInfo> .. </gmd:MD_Metadata>
Comment	–

A-II.2.2.Online digital transfer options

Metadata element name	Online digital transfer options
Definition	Information about online sources from which the resource can be obtained.
ISO 19115 number and name	277. onLine
ISO/TS 19139 path	distributionInfo/*/transferOptions/*/online/*/linkage
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	396. CI_OnlineResource
Domain	The following property is expected: - linkage (i.e. URL to the resource)
Implementing instructions	–
Example	-
Example XML encoding	<gmd:MD_Metadata .. <gmd:distributionInfo> <gmd:transferOptions>

	<gmd:MD_DigitalTransferOptions> <gmd:onLine> <gmd:CI_OnlineResource> <gmd:linkage> <gmd:URL>**http://wwwedisoft.pt/index**</ gmd:URL> </gmd:linkage> </gmd:CI_OnlineResource> </gmd:onLine> </gmd:MD_DigitalTransferOptions> </gmd:transferOptions> </gmd:distributionInfo> </gmd:MD_Metadata>
Comment	–

A-II.2.3.Topology level

Metadata element name	Topology level
Definition	Code which identifies the degree of complexity of the spatial relationships.
ISO 19115 number and name	177. topologyLevel
ISO/TS 19139 path	spatialRepresentationInfo /*/topologyLevel
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..1
Data type (and ISO 19115 no.)	MD_TopologyLevelCode
Domain	CodeList (See B.5.28 of ISO 19115)
Implementing instructions	–
Example	–
Example XML encoding	<gmd:MD_Metadata .. <gmd:spatialRepresentationInfo> <gmd:MD_VectorSpatialRepresentation> <gmd:topologyLevel> <gmd:MD_TopologyLevelCode codeList="**http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Codelist/ML_gmx/Codelists.xml#MD_TopologyLevelCode**" codeListValue="geometryOnly">**geometryOnly**</gmd:MD_TopologyLevelCode> > </gmd:topologyLevel> </gmd:MD_VectorSpatialRepresentation> </gmd:spatialRepresentationInfo> </gmd:MD_Metadata>
Comment	-

A-II.2.4.Source mapping scale

Metadata element name	Source mapping scale
Definition	Denominator of the representative fraction on a source map.
ISO 19115 number and name	94. scaleDenominator
ISO/TS 19139 path	dataQualityInfo/*/lineage/*/source/*/scaleDenominator
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..1
Data type (and ISO 19115 no.)	56. MD_RepresentativeFraction
Domain	Integer
Implementing instructions	Expression of two equivalent scales is not allowed in this metadata element.
Example	10000, 50000, 1000000
Example XML encoding	<gmd:MD_Metadata .. <gmd:dataQualityInfo> <gmd:lineage> <gmd:LI_Lineage> <gmd:source> <gmd:LI_Source> <gmd:scaleDenominator> <gmd:MD_RepresentativeFraction> <gmd:denominator> <gco:Integer>**50000**</gco:Integer> </gmd:denominator> </gmd:MD_RepresentativeFraction> </gmd:scaleDenominator> </gmd:LI_Source> </gmd:source> </gmd:LI_Lineage> </gmd:lineage> </gmd:dataQualityInfo> </gmd:MD_Metadata>
Comment	This metadata element contains only the denominator value, i.e. the whole expression of the scale like 1 : 50 000 is not allowed; this value has to be expressed as 50000.

A-II.2.5.Data Quality – Completeness - Omission

Metadata element name	Data Quality – Completeness - Omission
Definition	Data absent from the dataset, as described by the scope.
ISO 19115 number and name	18. dataQualityInfo
ISO/TS 19139 path	dataQualityInfo
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*

Data type (and ISO 19115 no.)	109 DQ_CompletenessOmission
Domain	Lines 100-107 from ISO 19115
Implementing instructions	This quality measure should answer the consumer question: How many real world items/instances does the content provider expect and how many of them are encoded at a given scope (spatial object type, dataset or dataset series)
Example	The following statement should e.g. Be expressed here correspondingly: 10 maritime datasets in the real world in the Mediterranean sea, 9 of them encoded in the dataset.
Example XML encoding	–
Comment	See Completeness / Omission section below for detailed information.

A-II.2.6.Data Quality – Positional accuracy – Absolute or external accuracy

Metadata element name	Data Quality – Positional accuracy – Absolute or external accuracy
Definition	Closeness of reported coordinate values to values accepted as being true.
ISO 19115 number and name	18. dataQualityInfo
ISO/TS 19139 path	dataQualityInfo
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	117. DQ_AbsoluteExternalPositionalAccuracy
Domain	Lines 100-107 from ISO 19115
Implementing instructions	–
Example	–
Example XML encoding	–
Comment	See Positional accuracy / Absolute or external accuracy section below for detailed information.

A-II.2.7.Data Quality – Thematic accuracy – Thematic classification correctness

Metadata element name	Data Quality – Thematic accuracy – Thematic classification correctness
Definition	Comparison of the classes assigned to features or their attributes to a universe of discourse.
ISO 19115 number and name	18. dataQualityInfo
ISO/TS 19139 path	dataQualityInfo
INSPIRE obligation / condition	Optional
INSPIRE multiplicity	0..*
Data type (and ISO 19115 no.)	125. DQ_ThematicClassificationCorrectness
Domain	Lines 100-107 from ISO 19115
Implementing instructions	This metadata should be filled, at least, with these elements: - valueUnite: UnitOfMeasure

	- value: Record	
Example	–	
Example XML encoding	–	
Comment	See Thematic Accuracy section below for detailed information.	

A-II.3. Data Quality

This section includes a description of data quality elements and sub-elements, as well as the associated basic data quality measures, to be used to describe data related to the spatial data theme ODYSSEA (see Table A-II.5).

Data quality information can be described at the level of spatial object (feature), spatial object type (feature type), dataset or dataset series. Data quality information at spatial object level is modelled directly in the application schema (i.e. data itself).

Aggregated data quality information should ideally be collected at the level of spatial object types and included in the dataset (series) metadata.

TABLE A-II.5: LIST OF ALL DATA QUALITY ELEMENTS USED IN THE ODYSSEA METADATA PROFILE

data quality list	Data quality element	Data quality sub-element	Scope(s)	Data quality scope
1	Completeness	Omission	dataset series; dataset; spatial object type	evaluation
2	Positional accuracy	Absolute or external accuracy	spatial object	evaluation
3	Thematic Accuracy	Classification correctness	dataset series; dataset	evaluation

A-II.3.1.Completeness

This data quality element enables the assessment of the presence of features, their attributes and relationships.

A-II.3.2.Omission

Omission should be documented using rate of missing items.

Name	Rate of missing items
Alternative name	–
Data quality element	Completeness
Data quality sub-element	Omission
Data quality basic measure	Error rate

Definition	Number of missing items in the dataset in relation to the number of items that should have been present.
Description	–
Parameter	–
Data quality value type	Real, percentage, ratio
Data quality value structure	–
Source reference	–
Example	0,0189 ; 98,11% ; 11:582
Measure identifier	7 (ISO 19138)

A-II.3.3.Positional accuracy: Absolute or external accuracy

Absolute or external accuracy should be documented using the mean value of positional uncertainties.

Name	Mean value of positional uncertainties (1D, 2D)
Alternative name	–
Data quality element	positional accuracy
Data quality subelement	absolute or external accuracy
Data quality basic measure	not applicable
Definition	Mean value of the positional uncertainties for a set of positions where the positional uncertainties are defined as the distance between a measured position and what is considered as the corresponding true position
Description	See ISO 19138
Parameter	–
Data quality value type	measure
Data quality value structure	–
Source reference	–
Example	–
Measure identifier	28 (ISO 19138)

A-II.3.4.Thematic Accuracy: Classification Correctness - Misclassification rate

This quality sub-element shows the comparison of the classes assigned to features or their attributes to a universe of discourse. The assessment of this quality sub-element should be stored in the metadata element DQ_ThematicClassificationCorrectness.

Name	Misclassification rate
Alternative name	–
Data quality element	Thematic accuracy
Data quality subelement	Classification correctness
Data quality basic measure	Error rate

Definition	Average number of incorrectly classified features in relation to the number of features that are supported to be within the dataset
Description	To be provided globaly as an average value for the whole dataset.
Parameter	–
Data quality value type	Real, percentage, ration
Data quality value structure	–
Source reference	–
Example	–
Measure identifier	61 (ISO 19138)

ANNEX III. Examples of data files structure

A-III.1. Example of ISO 19139 XML encoding

The XML source code written below represents an example of a full ODYSSEA metadata profile – i.e. all metadata elements according to the Commission Regulation No 1205/2008/EC (implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata) for spatial datasets and spatial dataset series as well as all ODYSSEA metadata elements described in the metadata profile documentation written above.

```
   <?xml version="1.0" encoding="utf-8" ?>

<gmd:MD_Metadata                                    xsi:schemaLocation="http://www.isotc211.org/2005/gmd
     http://www.isotc211.org/2005/gmd/metadataEntity.xsd"  xmlns:gmd="http://www.isotc211.org/2005/gmd"
     xmlns:gco="http://www.isotc211.org/2005/gco"          xmlns:gmx="http://www.isotc211.org/2005/gmx"
     xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:gml="http://www.opengis.net/gml">

<gmd:fileIdentifier>

     <gco:CharacterString>0de9b6b7-641c-441b-bf37-f03db0960cb8</gco:CharacterString>

    </gmd:fileIdentifier>

<gmd:language>

<gmd:LanguageCode

     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
    Schemas/resources/Codelist/ML_gmxCodelists.xml#LanguageCode"
    codeListValue="eng">eng</gmd:LanguageCode> </gmd:language>

<gmd:hierarchyLevel>

<gmd:MD_ScopeCode

     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
    Schemas/resources/Codelist/ML_gmxCodelists.xml#MD_ScopeCode"
    codeListValue="dataset">dataset</gmd:MD_ScopeCode> </gmd:hierarchyLevel>

<gmd:contact>

<gmd:CI_ResponsibleParty>

<gmd:organisationName>

    <gco:CharacterString>ODYSSEA Project</gco:CharacterString>

     </gmd:organisationName>

<gmd:contactInfo>

<gmd:CI_Contact>

<gmd:address>

<gmd:CI_Address>

<gmd:electronicMailAddress>

  <gco:CharacterString>ODYSSEA@ODYSSEA.eu</gco:CharacterString>
    </gmd:electronicMailAddress>

    </gmd:CI_Address>

    </gmd:address>

<gmd:onlineResource>

<gmd:CI_OnlineResource>

<gmd:linkage>
```

```
  <gmd:URL>http://ODYSSEAplatform.eu/</gmd:URL>
     </gmd:linkage>
     </gmd:CI_OnlineResource>
     </gmd:onlineResource>
     </gmd:CI_Contact>
     </gmd:contactInfo>


<gmd:role>
<gmd:CI_RoleCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_RoleCode"
     codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
       </gmd:role>
     </gmd:CI_ResponsibleParty>
     </gmd:contact>
<gmd:dateStamp>
  <gco:Date>2010-04-09</gco:Date>
     </gmd:dateStamp>
<gmd:spatialRepresentationInfo>
<gmd:MD_VectorSpatialRepresentation>
<gmd:topologyLevel>
<gmd:MD_TopologyLevelCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmx/Codelists.xml#MD_TopologyLevelCode"
     codeListValue="geometryOnly">geometryOnly</gmd:MD_TopologyLevelCode>
      </gmd:topologyLevel>
<gmd:geometricObjects>
<gmd:MD_GeometricObjects>
<gmd:geometricObjectType>
<gmd:MD_GeometricObjectTypeCode
      codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_Schemas/resources/Cod
      elist/ML_gmx/Codelists.xml#MD_GeometricObjectTypeCode"
      codeListValue="surface">surface</gmd:MD_GeometricObjectTypeCode> </gmd:geometricObjectType>
     </gmd:MD_GeometricObjects>
     </gmd:geometricObjects>
     </gmd:MD_VectorSpatialRepresentation>
     </gmd:spatialRepresentationInfo>
<gmd:referenceSystemInfo>
<gmd:MD_ReferenceSystem>
<gmd:referenceSystemIdentifier>
<gmd:RS_Identifier>
<gmd:code>
  <gco:CharacterString>WGS 1984</gco:CharacterString> </gmd:code>
<gmd:codeSpace>
```

```
  <gco:CharacterString>EPSG</gco:CharacterString>
     </gmd:codeSpace>
     </gmd:RS_Identifier>
     </gmd:referenceSystemIdentifier>
     </gmd:MD_ReferenceSystem>
     </gmd:referenceSystemInfo>
<gmd:identificationInfo>
<gmd:MD_DataIdentification>
<gmd:citation>
<gmd:CI_Citation>
<gmd:title>
        <gco:CharacterString>Sample ODYSSEA Dataset metadata</gco:CharacterString>
</gmd:title>
<gmd:date>
<gmd:CI_Date>
<gmd:date>
  <gco:Date>2018-01-01</gco:Date>
     </gmd:date>
<gmd:dateType>
<gmd:CI_DateTypeCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode"
     codeListValue="publication">publication</gmd:CI_DateTypeCode>
      </gmd:dateType>
     </gmd:CI_Date>
     </gmd:date>
<gmd:identifier>
<gmd:RS_Identifier>
<gmd:code>
  <gco:CharacterString>ODYSSEA_Xpto</gco:CharacterString> </gmd:code>
<gmd:codeSpace>
  <gco:CharacterString>MU</gco:CharacterString>
     </gmd:codeSpace>
     </gmd:RS_Identifier>
     </gmd:identifier>
     </gmd:CI_Citation>
     </gmd:citation>
<gmd:abstract>
   <gco:CharacterString>This dataset represents aggregated ODYSSEA data sources.</gco:CharacterString>
    </gmd:abstract>
<gmd:pointOfContact>
<gmd:CI_ResponsibleParty>
```

```
<gmd:organisationName>
    <gco:CharacterString>ODYSSEA Project</gco:CharacterString>
     </gmd:organisationName>
<gmd:contactInfo>
<gmd:CI_Contact>
<gmd:address>
<gmd:CI_Address>
<gmd:electronicMailAddress>
  <gco:CharacterString>ODYSSEA@ODYSSEA.eu</gco:CharacterString>
     </gmd:electronicMailAddress>
     </gmd:CI_Address>
     </gmd:address>
     </gmd:CI_Contact>
     </gmd:contactInfo>
<gmd:role>
<gmd:CI_RoleCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_RoleCode"
     codeListValue="processor">processor</gmd:CI_RoleCode> </gmd:role>
     </gmd:CI_ResponsibleParty>
     </gmd:pointOfContact>
<gmd:descriptiveKeywords>
<gmd:MD_Keywords>
<gmd:keyword>
  <gco:CharacterString>ODYSSEA</gco:CharacterString>
     </gmd:keyword>
<gmd:thesaurusName>
<gmd:CI_Citation>
<gmd:title>
  <gco:CharacterString>GEMET - INSPIRE themes, version 1.0</gco:CharacterString> </gmd:title>
<gmd:date>
<gmd:CI_Date>
<gmd:date>
  <gco:Date>2018-01-01</gco:Date>
     </gmd:date>
<gmd:dateType>
<gmd:CI_DateTypeCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode"
     codeListValue="publication">publication</gmd:CI_DateTypeCode>
     </gmd:dateType>
     </gmd:CI_Date>
     </gmd:date>
     </gmd:CI_Citation>
```

```
    </gmd:thesaurusName>
    </gmd:MD_Keywords>
    </gmd:descriptiveKeywords>
<gmd:resourceConstraints>
<gmd:MD_Constraints>
<gmd:useLimitation>
  <gco:CharacterString>no       conditions       apply</gco:CharacterString>
    </gmd:useLimitation>
    </gmd:MD_Constraints>
    </gmd:resourceConstraints>
<gmd:resourceConstraints>
<gmd:MD_LegalConstraints>
<gmd:accessConstraints>
<gmd:MD_RestrictionCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
    Schemas/resources/Codelist/ML_gmxCodelists.xml#MD_RestrictionCode"
    codeListValue="otherRestrictions">otherRestrictions</gmd:MD_RestrictionCode>
     </gmd:accessConstraints>
<gmd:otherConstraints>


  <gco:CharacterString>intellectual       property       rights</gco:CharacterString>
    </gmd:otherConstraints>
    </gmd:MD_LegalConstraints>
    </gmd:resourceConstraints>
<gmd:spatialResolution>
<gmd:MD_Resolution>
<gmd:equivalentScale>
<gmd:MD_RepresentativeFraction>
<gmd:denominator>
  <gco:Integer>50000</gco:Integer>
    </gmd:denominator>
    </gmd:MD_RepresentativeFraction>
    </gmd:equivalentScale>
    </gmd:MD_Resolution>
    </gmd:spatialResolution>
<gmd:spatialResolution>
<gmd:MD_Resolution>
  <gmd:distance gco:nilReason="missing" />
    </gmd:MD_Resolution>
    </gmd:spatialResolution>
<gmd:language>
<gmd:LanguageCode
```

```
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#LanguageCode"
     codeListValue="pt">pt</gmd:LanguageCode> </gmd:language>
<gmd:characterSet>
<gmd:MD_CharacterSetCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmx                    Codelists.xml#MD_ClassificationCode"
     codeListValue="utf8">utf8</gmd:MD_CharacterSetCode>
     </gmd:characterSet>
<gmd:topicCategory>
  <gmd:MD_TopicCategoryCode>geoscientificInformation</gmd:MD_TopicCategoryCode> </gmd:topicCategory>
<gmd:extent>
<gmd:EX_Extent>
<gmd:geographicElement>
<gmd:EX_GeographicBoundingBox>
<gmd:westBoundLongitude>
  <gco:Decimal>12.09</gco:Decimal>
     </gmd:westBoundLongitude>
<gmd:eastBoundLongitude>
  <gco:Decimal>18.85</gco:Decimal>
     </gmd:eastBoundLongitude>
<gmd:southBoundLatitude>
<gco:Decimal>48.58</gco:Decimal>
     </gmd:southBoundLatitude>
<gmd:northBoundLatitude>
  <gco:Decimal>51.05</gco:Decimal>
     </gmd:northBoundLatitude>
     </gmd:EX_GeographicBoundingBox>
     </gmd:geographicElement>
<gmd:temporalElement>
<gmd:EX_TemporalExtent>
<gmd:extent>
<gml:TimePeriod gml:id="ID6w561g20-4vwp-8l1i-6d5l-hhu5sv1h2fw1" xsi:type="gml:TimePeriodType">
  <gml:beginPosition />
  <gml:endPosition />
     </gml:TimePeriod>
       </gmd:extent>
     </gmd:EX_TemporalExtent>
     </gmd:temporalElement>
     </gmd:EX_Extent>
     </gmd:extent>
     </gmd:MD_DataIdentification>
     </gmd:identificationInfo>
```

```
<gmd:distributionInfo>
<gmd:MD_Distribution>
<gmd:distributionFormat>
<gmd:MD_Format>
<gmd:name>
  <gco:CharacterString>3.2.1</gco:CharacterString>
    </gmd:name>
<gmd:version>
  <gco:CharacterString>GML</gco:CharacterString>
    </gmd:version>
    </gmd:MD_Format>
    </gmd:distributionFormat>
<gmd:transferOptions>
<gmd:MD_DigitalTransferOptions>
<gmd:onLine>
<gmd:CI_OnlineResource>
<gmd:linkage>
  <gmd:URL>http://ODYSSEAplatform.eu/</gmd:URL> </gmd:linkage>
    </gmd:CI_OnlineResource>
    </gmd:onLine>
    </gmd:MD_DigitalTransferOptions>
    </gmd:transferOptions>
    </gmd:MD_Distribution>
    </gmd:distributionInfo>
<gmd:dataQualityInfo>
<gmd:DQ_DataQuality>
<gmd:scope>
<gmd:DQ_Scope>
<gmd:level>
<gmd:MD_ScopeCode        codeListValue="dataset"
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
     Schemas/resources/Codelist/ML_gmxCodelists.xml#MD_ScopeCode">dataset</
     gmd:MD_ScopeCode>
    </gmd:level>
    </gmd:DQ_Scope>
    </gmd:scope>
<gmd:report>
<gmd:DQ_DomainConsistency xsi:type="gmd:DQ_DomainConsistency_Type"> <gmd:result>
<gmd:DQ_ConformanceResult xsi:type="gmd:DQ_ConformanceResult_Type">
<gmd:specification>
<gmd:CI_Citation>
<gmd:title>
```

```
<gco:CharacterString>ODYSSEA dataset</gco:CharacterString>
</gmd:title>
<gmd:date>
<gmd:CI_Date>
<gmd:date>
  <gco:Date>2018-01-01</gco:Date>
    </gmd:date>
    <gmd:dateType>
    <gmd:CI_DateTypeCode
     codeList="http://standards.iso.org/ittf/PubliclyAvailableStandards/ISO_19139_
    Schemas/resources/Codelist/ML_gmxCodelists.xml#CI_DateTypeCode"
    codeListValue="revision">revision</gmd:CI_DateTypeCode> </gmd:dateType>
    </gmd:CI_Date>
    </gmd:date>
    </gmd:CI_Citation>
    </gmd:specification>
<gmd:explanation>
        <gco:CharacterString>See the Commission Regulation No 1205/2008/EC.</gco:CharacterString>
    </gmd:explanation>
<gmd:pass>
  <gco:Boolean>true</gco:Boolean>
    </gmd:pass>
    </gmd:DQ_ConformanceResult>
    </gmd:result>
    </gmd:DQ_DomainConsistency>
    </gmd:report>
<gmd:report>
<gmd:DQ_CompletenessOmission>
<gmd:result>
<gmd:DQ_QuantitativeResult>
<gmd:valueType>
  <gco:RecordType>percentage</gco:RecordType>
</gmd:valueType>
<gmd:valueUnit>
<gml:BaseUnit gml:id="perc">
  <gml:identifier codeSpace="" />
  <gml:unitsSystem />
    </gml:BaseUnit>
    </gmd:valueUnit>
<gmd:value>
  <gco:Record>98,4%</gco:Record>
    </gmd:value>
```

```
    </gmd:DQ_QuantitativeResult>
    </gmd:result>
    </gmd:DQ_CompletenessOmission>
    </gmd:report>
<gmd:report>
<gmd:DQ_AbsoluteExternalPositionalAccuracy>
<gmd:result>
<gmd:DQ_QuantitativeResult>
<gmd:valueType>
  <gco:RecordType>measure</gco:RecordType>
    </gmd:valueType>
<gmd:valueUnit>
<gml:BaseUnit gml:id="meters">
  <gml:identifier codeSpace="" />
  <gml:unitsSystem />
    </gml:BaseUnit>
    </gmd:valueUnit>
<gmd:value>
  <gco:Record>1,5m</gco:Record>
    </gmd:value>
    </gmd:DQ_QuantitativeResult>
    </gmd:result>
    </gmd:DQ_AbsoluteExternalPositionalAccuracy>
    </gmd:report>
<gmd:report>
<gmd:DQ_ThematicClassificationCorrectness>
<gmd:result>
<gmd:DQ_QuantitativeResult>
<gmd:valueType>
  <gco:RecordType>percentage</gco:RecordType>
    </gmd:valueType>
<gmd:valueUnit>
<gml:BaseUnit gml:id="years">
  <gml:identifier codeSpace="" />
  <gml:unitsSystem />
    </gml:BaseUnit>
    </gmd:valueUnit>
<gmd:value>
  <gco:Record>2,4%</gco:Record>
    </gmd:value>
    </gmd:DQ_QuantitativeResult>
    </gmd:result>
```

```
    </gmd:DQ_ThematicClassificationCorrectness>
    </gmd:report>
<gmd:lineage>
<gmd:LI_Lineage>
<gmd:statement>
  <gco:CharacterString>This dataset has been created while using several underlying
    datasets..</gco:CharacterString>
    </gmd:statement>
<gmd:source>
<gmd:LI_Source>
<gmd:scaleDenominator>
<gmd:MD_RepresentativeFraction>
<gmd:denominator>
  <gco:Integer>50000</gco:Integer>
    </gmd:denominator>
    </gmd:MD_RepresentativeFraction>
    </gmd:scaleDenominator>
<gmd:sourceCitation>
<gmd:CI_Citation>
<gmd:title>
       <gco:CharacterString>ODYSSEA Data</gco:CharacterString>
    </gmd:title>
<gmd:date>
<gmd:CI_Date>
<gmd:date>
  <gco:Date>2018-01-01</gco:Date>
    </gmd:date>
<gmd:dateType>
<gmd:CI_DateTypeCode
     codeList="http://www.isotc211.org/2005/resources/codeList.xml#CI_DateType
     Code" codeListValue="creation" />
    </gmd:dateType>
    </gmd:CI_Date>
    </gmd:date>
</gmd:CI_Citation>
</gmd:sourceCitation>
</gmd:LI_Source>
</gmd:source>
</gmd:LI_Lineage>
</gmd:lineage>
</gmd:DQ_DataQuality>
</gmd:dataQualityInfo>
```

```
</gmd:MD_Metadata>
```

A-III.2. Example of CMEMS data file

CMEMS uses netCDF format for the data files that it is handling. An example of the CMEMS netCDF files header is presented below:

```
NetCDF-3 Classic global-analysis-forecast-phy-001-024_1506677952654.nc {
dimensions:
        time = 2 ;
        depth = 1 ;
        latitude = 361 ;
        longitude = 61 ;
variables:
        // Preference 'PRESERVE_FVD':  false,
        // dimensions consistent with ncBrowse, not with native MATLAB netcdf package.
        single time(time), shape = [2]
                time:long_name = "Time (hours since 1950-01-01)" ;
                time:standard_name = "time" ;
                time:calendar = "gregorian" ;
                time:units = "hours since 1950-01-01 00:00:00" ;
                time:axis = "T" ;
                time:_CoordinateAxisType = "Time" ;
        int16 thetao(time,depth,latitude,longitude), shape = [2 1 361 61]
                thetao:_CoordinateAxes = "time depth latitude longitude " ;
                thetao:long_name = "Temperature" ;
                thetao:standard_name = "sea_water_potential_temperature" ;
                thetao:units = "degrees_C" ;
                thetao:unit_long = "Degrees Celsius" ;
                thetao:_FillValue = -32767 s;
                thetao:add_offset = 21 ;
                thetao:scale_factor = 0.000732444 ;
                thetao:cell_methods = "area: mean" ;
        single longitude(longitude), shape = [61]
                longitude:step = 0.083328 f;
                longitude:units = "degrees_east" ;
                longitude:unit_long = "Degrees East" ;
                longitude:long_name = "Longitude" ;
                longitude:standard_name = "longitude" ;
                longitude:axis = "X" ;
                longitude:_CoordinateAxisType = "Lon" ;
        single latitude(latitude), shape = [361]
                latitude:step = 0.083336 f;
```

latitude:units = "degrees_north" ;
latitude:unit_long = "Degrees North" ;
latitude:long_name = "Latitude" ;
latitude:standard_name = "latitude" ;
latitude:axis = "Y" ;
latitude:_CoordinateAxisType = "Lat" ;
single depth(depth), shape = [1]
depth:units = "m" ;
depth:positive = "down" ;
depth:unit_long = "Meters" ;
depth:long_name = "Depth" ;
depth:standard_name = "depth" ;
depth:axis = "Z" ;
depth:_CoordinateAxisType = "Height" ;
depth:_CoordinateZisPositive = "down" ;

//global attributes:
:title = "daily mean fields from Global Ocean Physics Analysis and Forecast updated Daily" ;
:institution = "MERCATOR OCEAN" ;
:references = "http://www.mercator-ocean.fr" ;
:source = "MERCATOR PSY4QV3R1" ;
Conventions = "CF-1.0" ;
history = "Data extracted from dataset http://opendap-glo.mercator-ocean.fr:8080/thredds/dodsC/global-analysis-forecast-phy-001-024" ;
:time_min = 594036 ;
:time_max = 594060 ;
:julian_day_unit = "hours since 1950-01-01 00:00:00" ;
:z_min = 0.494025 ;
:z_max = 0.494025 ;
:latitude_min = 20 ;
:latitude_max = 50 ;
:longitude_min = 0 ;
:longitude_max = 5 ;

ANNEX IV. Data Privacy Procedures

A-IV.1. Data protection principles

1. The data privacy aims to ensure compliance with the Act. The Act sets out eight principles with which any party handling personal data must comply. All personal data:

2. Must be processed fairly and lawfully, meaning that at least one of the following conditions must be met:
   - The data subject has given his or her consent to the processing;
   - The processing is necessary for the performance of a contract to which the data subject is a party, or for the taking of steps at the request of the data subject with a view to entering into a contract;
   - The processing is necessary for compliance with any legal obligation to which the data controller is subject, other than an obligation imposed by contract;
   - The processing is necessary in order to protect the vital interests of the data subject;
   - The processing is necessary for the administration of justice, for the exercise of any functions of either House of Parliament, for the exercise of any functions conferred on any person by or under any enactment, for the exercise of any functions of the Crown, a Minister of the Crown or a government department, or for the exercise of any other functions of a public nature exercised in the public interest by any person;
   - The processing is necessary for the purposes of legitimate interests pursued by the data controller or by the third party or parties to whom the data is disclosed, except where the processing is unwarranted in any particular case by reason of prejudice to the rights and freedoms or legitimate interests of the data subject.

3. Where the personal data is sensitive personal data (defined below in Part 4 of this Policy), at least one of the following conditions must be met:
   - The data subject has given his or her explicit consent to the processing of the personal data;
   - The processing is necessary for the purposes of exercising or performing any right or obligation which is conferred or imposed by law on the data controller in connection with employment;
   - The processing is necessary in order to protect the vital interests of the data subject or another person in a case where consent cannot be given by or on behalf of the data subject, or the data controller cannot reasonably be expected to obtain the consent of the data subject, or in order to protect the vital interests of another person, in a case where consent by or on behalf of the data subject has been unreasonably withheld;
   - The processing is carried out in the course of the legitimate activities of any body or association which is not established or conducted for profit, and exists for political, philosophical, religious or trade-union purposes, is carried out with appropriate safeguards for the rights and freedoms of data subjects, relates only to individuals who either are members of the body or association or have regular contact with it in connection with its purposes, and does not involve disclosure of the personal data to a third party without the consent of the data subject;

- The information contained in the personal data has been made public as a result of steps deliberately taken by the data subject;
- The processing is necessary for the purpose of, or in connection with, any legal proceedings (including prospective legal proceedings), the processing is necessary for the purpose of obtaining legal advice, or is otherwise necessary for the purposes of establishing, exercising or defending legal rights;
- The processing is necessary for the administration of justice, for the exercise of any functions conferred on any person by or under an enactment, or for the exercise of any functions of the Crown, a minister of the Crown or a government department;
- The processing is either the disclosure of sensitive personal data by a person as a member of an anti-fraud organisation or otherwise in accordance with any arrangements made by such an organisation, or any other processing by that person or another person of sensitive personal data so disclosed, and is necessary for the purposes of preventing fraud or a particular kind of fraud;
- The processing is necessary for medical purposes and is undertaken by a health professional, or a person who in the circumstances owes a duty of confidentiality which is equivalent to that which would arise if that person were a health professional;
- The processing is of sensitive personal data consisting of information as to racial or ethnic origin, the processing is necessary for the purpose of identifying or keeping under review the existence or absence of equality of opportunity or treatment between persons of different racial or ethnic origins, with a view to enabling such equality to be promoted or maintained, and is carried out with appropriate safeguards for the rights and freedoms of data subjects.

4. Must be obtained only for specified and lawful purposes and shall not be processed in any manner which is incompatible with those purposes;
5. Must be adequate, relevant and not excessive with respect to the purposes for which it is processed;
6. Must be accurate and, where appropriate, kept up to date;
7. Must be kept for no longer than is necessary in light of the purpose(s) for which it is processed;
8. Must be processed in accordance with the rights of data subjects under the Act (for which, see Part 3 of this Policy);
9. Must be protected against unauthorised or unlawful processing, accidental loss, destruction or damage through appropriate technical and organisational measures; and
10. Must not be transferred to a country or territory outside of the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects in relation to the processing of personal data.

A-IV.2. Rights of Data Subjects

Under the Act, data subjects have the following rights:

- The right to access a copy of their personal data held by ODYSSEA by means of a Subject Access Request (for which, see Part 8 of this Policy);

- The right to object to any processing of his or her personal data that is likely to cause (or that is causing) damage or distress;
- The right to prevent processing for direct marketing purposes;
- The right to object to decisions being taken by automated means (where such decisions will have a significant effect on the data subject) and to be informed when any such decision is taken (in which case the data subject has the right to require the data controller (by written notice) to reconsider the decision;
- The right to have inaccurate personal data rectified, blocked, erased or destroyed in certain circumstances;
- The right to claim compensation for damage caused by the Company's breach of the Act.

A-IV.3. Personal Data

Personal data is defined by the Act as data which relates to a living individual who can be identified from that data or from that data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

The Act also defines "sensitive personal data" as personal data relating to the racial or ethnic origin of the data subject; their political opinions; their religious (or similar) beliefs; trade union membership; their physical or mental health condition; their sexual life; the commission or alleged commission by them of any offence; or any proceedings for any offence committed or alleged to have been committed by them, the disposal of such proceedings or the sentence of any court in such proceedings.

ODYSSEA only holds personal data that is directly relevant to its dealings with a given data subject. That data will be collected, held, and processed in accordance with the data protection principles and with this Policy. The following data may be collected, held and processed by ODYSSEA:

- Name, Address, Phone Numbers and Email Addresses;
- Area of work, main requirements of the users' activity that might be fulfilled by the offered services;
- Any other information provided to us by users.

A-IV.4. Processing Personal Data

Any and all personal data collected by ODYSSEA is collected in order to ensure that ODYSSEA can provide the best possible service to its customers.

Certain data collected by ODYSSEA, such as IP addresses, certain information gathered by cookies, pseudonyms and other non-identifying information will nonetheless be collected, held and processed to the same standards as personal data. In particular, ODYSSEA shall ensure that:

- All personal data collected and processed is collected and processed fairly and lawfully;

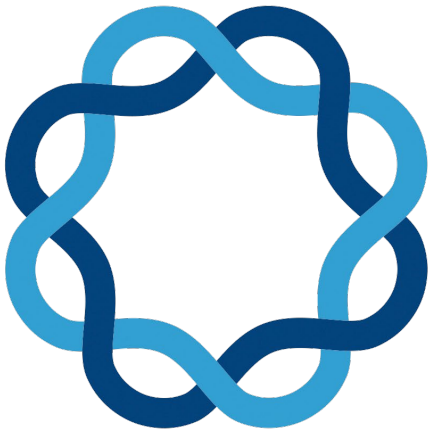
- Data subjects are always made fully aware of the reasons for the collection of personal data and are given details of the purpose(s) for which the data will be used;
- Personal data is only collected to the extent that is necessary to fulfil the purpose(s) for which it is required;
- All personal data is accurate at the time of collection and kept accurate and up to date while it is being held and/or processed;
- No personal data is held for any longer than necessary in light of the purpose(s) for which it is required;
- A suitable online privacy policy is implemented, maintained and followed;
- Whenever cookies or similar technologies are used online, they shall be used strictly in accordance with the requirements of the Privacy and Electronic Communications Regulations, providing full details of cookie use and guidance on privacy;
- Individuals are provided with a simple, accessible method of amending any data submitted by them online;
- All personal data is held in a safe and secure manner taking all appropriate technical and organisational measures to protect the data.

A-IV.5. Data Protection Procedures

ODYSSEA shall ensure that all of its employees comply with the following when working with personal data:

- All emails containing personal data must be sent securely;
- If personal data is being viewed on a computer screen and the computer in question is to be left unattended for any period of time, the user must lock the computer and screen before leaving it;
- No personal data should be transferred to any device personally belonging to an employee and personal data may only be transferred to devices belonging to agents, contractors, or other parties working on behalf of ODYSSEA where the party in question has agreed to comply fully with the letter and spirit of this Policy and of the Act (which may include demonstrating to the Company that all suitable technical and organisational measures have been taken);
- All personal data stored electronically should be backed up regularly with backups stored [onsite] AND/OR [offsite].
- All electronic copies of personal data should be stored securely using passwords and data encryption;
- All passwords used to protect personal data should be changed regularly and should not use words or phrases that can be easily guessed or otherwise compromised. All passwords must contain a combination of uppercase and lowercase letters, numbers, and symbols [. All software used by the Company is designed to require such passwords];

- Under no circumstances should any passwords be written down or shared between any employees irrespective of seniority or department. If a password is forgotten, it must be reset using the applicable method. IT staff do not have access to passwords;
- All personal data shall be regularly reviewed for accuracy and completeness. Where ODYSSEA has regular contact with data subjects, any personal data held about those data subjects should be confirmed at least annually. If any personal data is found to be out of date or otherwise inaccurate, it should be updated and/or corrected immediately where possible. If any personal data is no longer required, it should be securely deleted and disposed of;
- Where personal data held by ODYSSEA is used for marketing purposes, it shall be the responsibility of the Managing Director to ensure that no data subjects have added their details to any marketing preference databases including, but not limited to, the Telephone Preference Service, the Mail Preference Service, the Email Preference Service, and the Fax Preference Service. Such details should be checked at least annually.



ODYSSEA

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727277. The report reflects only authors' views and the Commission is not responsible for any use that may be made of the information it contains.