HiSea Deliverable 3.2

Data Management Plan

Work Package Number: 3

Work Package Title: Service Specifications and User Requirements

HISEA Project Information	
Project full title	High Resolution Copernicus-Based Information Services at Sea for Ports and Aquaculture
Project acronym	HISEA
Grant agreement number	821934
Project coordinator	Dr. Ghada El Serafy
Project start date and duration	1st January, 2019, 30 months
Project website	https://hiseaproject.com/

Deliverable Information	
Work package number	3
Work package title	Service Specifications and User Requirements
Deliverable number	3.2
Deliverable title	Data Management Plan
Description	First set of guidelines related with HiSea data management
Lead beneficiary	HIDROMOD
Lead Author(s)	Adélio Silva (Hidromod)
Contributor(s)	
Revision number	V0.1
Revision Date	29/06/2019
Status (Final (F), Draft (D), Revised Draft (RV))	F

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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))	RE

Document History			
Revision	**Date**	**Modification**	**Author**
0.1	09/05/2019	Text Updating	Sandra Gaytan
0.2	24/05/2019	Text Updating	Anna Spinosa
0.3	14/06/2019	Text Updating	Adélio Silva
0.4	14/06/2019	Text Updating	Sandra Gaytan

Approvals				
	Name	**Organisation**	**Date**	**Signature (initials)**
Coordinator	Ghada El Serafy	Deltares		
WP Leaders	Adélio Silva	HIDROMOD		

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

Table of Contents

1 Introduction 9

2 Data standards to be adopted in HiSea 11

2.1 Data quality control 11

2.2 In situ observations quality control 12

2.3 Forecasts quality control 13

3 Data integration and fusion 14

4 Data management 15

4.1 Data vocabulary 15

4.2 Metadata 15

4.3 Metadata Catalogue Service 17

4.3.1 Catalogue Service for Web (CSW) 18

4.3.2 Harvesting 18

4.4 Guidelines on using metadata elements 19

4.4.1 Lineage 19

4.4.2 Temporal reference 19

4.4.3 Topic category 19

4.4.4 Keyword 19

5 HiSea datasets 20

5.1 Dataset-level metadata 20

5.2 Data format standards 21

5.2.1 Ocean Data View data model and netCDF Format 21

6 Data privacy policy 23

6.1 General principles 23

6.2 Use of Cookies 24

7 References 25

ANNEX I: Mandatory and optional metadata elements 27

Mandatory metadata elements 27

Optional metadata elements 33

Data Quality Info 38

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

ANNEX II: Examples of data files structure _____ *41*

Example of ISO 19139 XML encoding _____ 41

Example of CMEMS data file _____ 48

ANNEX III: Data Privacy Procedures _____ *50*

Data protection principles _____ 50

Rights of Data Subjects _____ 52

Personal Data _____ 52

Processing Personal Data _____ 52

Data Protection Procedures _____ 53

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

List of Tables

Table 1: Metadata for spatial datasets and spatial dataset series specified in the INSPIRE Metadata Regulation [REGULATION 1205/2008/EC] 27

Table 2: Mandatory and conditional theme-specific metadata for the HiSea metadata profile.....28

Table 3: Optional theme-specific metadata for the HiSea metadata profile.....33

Table 4: List of all data quality elements used in the HiSea metadata profile.....38

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

Executive Summary

This deliverable will provide the guidelines and the first version of the Data Management Plan (DMP) for the HiSea Platform including strategies for improving data management, data privacy and data quality control. As an EU funded project, the data management will tune in to specific aspects of data management within the European context such as existing networks of data and requirements of European industry and other end-users.

According to the requirements of H2020 Programme (H2020, 2016), *Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable (FAIR), a DMP should include information on:*

- *the handling of research data during and after the end of the project;*
- *what data will be collected, processed and/or generated;*
- *which methodology and standards will be applied;*
- *whether data will be shared/made open access and;*
- *how data will be curated and preserved (including after the end of the project).*

The HiSea services will be supported by a diversity of data sets from existing networks and platforms (CMEMS, SEADATANET, EMODNET, NOAA, etc.) and from HiSea produced data (operational numerical models' data, in-situ data, etc.) that should be ready to answer to the above referred requirements.

Some of these data sets, because daily acquired from external providers do not need these concerns since they will not be stored in HiSea platform.

However, it is expectable that some local existing data sets will be also collected and used to produce, validate or complement the provided services (and may be useful to other external users). The latest should follow these requirements. In this context, the issues of data management and planning still represent a central concern to HiSea.

This document is a first version of the DMP and summarizes a set of guidelines that may have an impact already in the first stage of the project where the platform architecture is being defined and the data management procedures are being agreed. These guidelines are mostly focused on the best practices 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 integrated access to data in order to facilitate the integration in 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 (WMS, WFS, etc.) to facilitate development; and facilitate the visibility of existing data and the identification of gaps.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

Other relevant issues such as those related with what data will be collected, processed and/or generated, how data will be curated and preserved or whether data will be shared/made open access will require a deeper learning about the data that will be available/produced, eventual existence of restrictions to dissemination from data acquired by the users or even the technologic characteristics/constraints of the infrastructure where the data will be stored. A proper perception of these issues will become clear as the project goes on.

An updated version of the DMP will be delivered in later in the project in order to incorporate the implementation process, lessons learned, and barriers overcome in data management while deploying the HiSea Platform.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

1 Introduction

HiSea aims to provide a set of services focused on different coastal user's needs (navigation safety, ports operations, aquacultures, 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 first agreed strategies to be implemented to improve data management and planning, data privacy and data quality control of HiSea services. According the EU recommendations these strategies will take in consideration the data that will be collected, processed and/or generated, how this data will be handled during and after the project and the standards that will be adopted following the FAIR (*findable, accessible, interoperable and re-usable*) data concept.

In order to make data findable, it is required to assure that the data produced and/or used in the project is discoverable with metadata, identifiable and locatable by means of a standard identification mechanism. Concerning the standards, special attention will be devoted to the following 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 following widely accepted standards also contributes to reduce the duplication of efforts among agencies; to improve the quality and reduce costs related to geospatial information, thus making oceanographic data more accessible to the public and helping to establish key partnerships to increase data availability. Aiming to contribute to these objectives, HiSea will adopt the procedures already proposed in 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 (like CMEMS) or static data sets (like EMODnet) will follow procedures like 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.

The data integration and fusion policies to adopt in HiSea is another relevant issue 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, to build the best available datasets or fuse different data sources to produce aggregated data. Although not being an easy ground, a proper address of this issue may represent a valuable contribution to improve data accuracy and robustness of models' initial and boundary conditions and to provide to the users comprehensive data that merge together different data sets based on reliable criteria.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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 HiSea. 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 daily to the end users.

In relation with data privacy (data protection and the rights of platform end-users, customers and business contacts), it is apparent that HiSea will assure the respect of their 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 of the state of the art and the procedures to be adopted in HiSea will be provided. Note that at this stage this document mostly focuses in defining the guidelines to be followed throughout HiSea platform development, and although it does not reflect yet a practical implementation of these guidelines which will be subject of a later document.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

2 Data standards to be adopted in HiSea

2.1 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, from which part is reproduced bellow, 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 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 can 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.

HiSea services data flux will be managed automatically by the HiSea 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. Regarding 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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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 using 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).

A guideline of recommended QC procedures has been compiled by project SeaDataNet 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 HiSea.

2.2 In situ observations quality control

The quality control of observations may 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. At the second phase a tool may be run to periodically 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 with this aim.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

2.3 Forecasts quality control

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

3 Data integration and fusion

The issue of the best strategies to adopt when it comes to merging datasets obtained from different data sources, to build the best available datasets or fuse different data sources to produce aggregated data, indices and products is not an easy ground. A possible solution when we have different solutions with different resolutions for the same area is to make a fusion of these data and offer a unique integrated dataset. Another option is to provide all datasets separately with an eventual option of an integrated solution. No matter the adopted solution, the final objective of the data integration and fusion is to contribute to improve data accuracy and robustness of models' initial and boundary conditions and to provide to the users comprehensive data that merge together different data sets based on reliable criteria.

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

4 Data management

This document summarizes the concept of metadata that is intended to be adopted by HiSea 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.

4.1 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 terms of reference, the problem of ambiguities related to the 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.

HiSea 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[1].

This new vocabulary is mapped to the standards recommended for HiSea 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)[2].

4.2 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 can be defined as the **data about the data**. Metadata describes how, when and by whom a 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, ans/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 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?);

[1] https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/A05/
[2] http://www.marinespecies.org/aphia.php?p=webservice

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

- 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 the 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. 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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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 Annex 1 of this document.

4.3 Metadata Catalogue Service

A **Metadata Catalogue Service** is a mechanism for storing and accessing descriptive metadata which 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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.

4.3.1 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*".

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 HIS HiSea EA, the HiSea ISO metadata profile will be developed and used as described in this document's metadata sections.

The diagram presented below illustrates a generic view of the CSW protocol and architecture.

4.3.2 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.

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).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

4.4 Guidelines on using metadata elements

4.4.1 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).

4.4.2 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.

4.4.3 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.

4.4.4 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

5 HiSea datasets

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

The structure described in this document is compliant to the existing ISO standards for metadata – 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 deliverable is to develop a metadata profile for HiSea geographic datasets and time-series datasets, within the framework of these ISO standards, to support the interoperability between the different metadata and/or GIS platforms. The metadata model to be adopted in HiSea is described in more detail in Annex I.

5.1 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[3] 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:

- first column provides a reference to a more detailed description;
- second column specifies the name of the metadata element;
- third column specifies the multiplicity;
- fourth column specifies the condition, under which the given element becomes mandatory (only for the first and second tables).

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

5.2 Data format standards

5.2.1 Ocean Data View data model and netCDF Format

As part of the HiSea services, data sets will be accessible via download services. Delivery of data to users requires common data transport 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.

The CF metadata conventions (http://cf-pcmdi.llnl.gov/) are designed to promote the processing and sharing of 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 based on 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*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

'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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

6 Data privacy policy

6.1 General principles

Basic principles regulated by the data protection Act will be observed namely:

- ✓ HiSea only hold the necessary personal data 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 hold for as long as necessary. Once data are no longer needed it will be deleted from HiSea records by the HiSea 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.
- ✓ HiSea 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 HiSea 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. HiSea will never sell, rent or exchange mailing lists of personal data.
- ✓ All HiSea 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.
- ✓ HiSea 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, HiSea never send bulk unsolicited emails, (popularly known as Spams) to any email addresses.
- ✓ HiSea may send emails to existing end-users and customers or prospective end-users and customers having enquired or registered in HiSea platform, regarding products or services directly provided by HiSea platform.
- ✓ All emails sent by HiSea will be clearly marked as originating from this platform. All such emails will also include clear instructions on how to unsubscribe from HiSea 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

6.2 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 providing information to the owner of the site.

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

HiSea may use Google Analytics to collect quantitative information on platform's performance and end-users' interaction with the platform. HiSea 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

7 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) No/... 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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

ANNEX I: 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 1*) as well as the HiSea -specific mandatory metadata elements specified in Table 2 and optionally elements specified in Table 3.

Mandatory metadata elements

Table 1: 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 a resolution 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..*	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

10.2	Metadata date	1	
10.3	Metadata language	1	

Table 2: Mandatory and conditional theme-specific metadata for the HiSea metadata profile

HiSea **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

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

	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).

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 HISEA project: Default Encoding name: HISEA version: version 3.0; GML, version 3.2.1
Implementing instructions	–
Example	name: HISEA 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>

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

	</gmd:MD_Distribution> </gmd:distributionInfo> .. </gmd:MD_Metadata>
Comment	–

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	-

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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, HISEA 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>**HISEA 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	–

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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	–

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

Optional metadata elements

The metadata describing a spatial dataset or a spatial dataset series related to HiSea should comprise the HiSea - specific metadata elements specified in Table 3. This table contains all metadata elements that have a multiplicity of 0..1 or 0..*.

Table 3: Optional theme-specific metadata for the HiSea metadata profile

HiSea **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..*

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_gmxCodelists.xml#MD_SpatialRepresentationType Code"

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

	codeListValue="vector">vector</gmd:MD_SpatialRepresentationTypeCode> </gmd:spatialRepresentationType> .. </gmd:identificationInfo> .. </gmd:MD_Metadata>
Comment	–

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	–

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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	-

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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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.

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)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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.

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.

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	–

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

Example XML encoding	–	
Comment	See Thematic Accuracy section below for detailed information.	

Data Quality Info

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 HiSea (see Table 4).

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 4: List of all data quality elements used in the HiSea metadata profile

Section of the HISEA 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

Completeness

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

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	–

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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)

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)

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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

ANNEX II: Examples of data files structure

Example of ISO 19139 XML encoding

The XML source code written below represents an example of a full HiSea 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 HiSea 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>HISEA Project</gco:CharacterString>
       </gmd:organisationName>
<gmd:contactInfo>
<gmd:CI_Contact>
<gmd:address>
<gmd:CI_Address>
<gmd:electronicMailAddress>
   <gco:CharacterString>HISEA@HISEA.eu</gco:CharacterString> </gmd:electronicMailAddress>
       </gmd:CI_Address>
       </gmd:address>
<gmd:onlineResource>
<gmd:CI_OnlineResource>
<gmd:linkage>
   <gmd:URL>http://HISEAplatform.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>
```

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
        </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/Codelist/ML_g
         mx/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 HISEA 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>HISEA_Xpto</gco:CharacterString> </gmd:code>
<gmd:codeSpace>
   <gco:CharacterString>MU</gco:CharacterString>
        </gmd:codeSpace>
```

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
        </gmd:RS_Identifier>
        </gmd:identifier>
        </gmd:CI_Citation>
        </gmd:citation>
<gmd:abstract>
    <gco:CharacterString>This dataset represents aggregated HISEA data sources.</gco:CharacterString>
     </gmd:abstract>
<gmd:pointOfContact>
<gmd:CI_ResponsibleParty>
<gmd:organisationName>
      <gco:CharacterString>HISEA Project</gco:CharacterString>
       </gmd:organisationName>
<gmd:contactInfo>
<gmd:CI_Contact>
<gmd:address>
<gmd:CI_Address>
<gmd:electronicMailAddress>
   <gco:CharacterString>HISEA@HISEA.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>HISEA</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>
```


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
<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 />
```

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
      <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://HISEAplatform.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>HISEA 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>
```


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
        </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="" />
```


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
    <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>HISEA 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>
```


This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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" ;
```

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

```
                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 ;
```

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

ANNEX III: Data Privacy Procedures

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 anybody 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

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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 the necessary time 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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 HiSea 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.

Personal Data

Personal data are 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.

HiSea only holds personal data that are 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 HiSea:

- 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.

Processing Personal Data

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

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

Certain data collected by HiSea, 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, HiSea shall ensure that:

- All personal data are 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 are only collected to the extent that is necessary to fulfil the purpose(s) for which it is required;
- All personal data are accurate at the time of collection and kept accurate and up to date while they are being held and/or processed;
- No personal data are held for any longer than necessary in light of the purpose(s) for which they are 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 are held in a safe and secure manner taking all appropriate technical and organisational measures to protect the data.

Data Protection Procedures

HiSea 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 are 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 HiSea 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].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934

- 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 HiSea has regular contact with data subjects, any personal data held about those data subjects should be confirmed at least annually. If any personal data are found to be out of date or otherwise inaccurate, they should be updated and/or corrected immediately where possible. If any personal data are no longer required, they should be securely deleted and disposed of;
- When personal data held by HiSea are 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.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 821934