EMODnet Data Ingestion and safe-keeping of marine data

Submission Form - Workflow procedure and Metadata Elements

Version: V1.1.0
Date: 5 February 2017
Prepared by: HCMR, MARIS and GET (subcontractor of HCMR) with input of consortium partners
# Table of Contents

Table of Contents

1. Introduction ................................................................................................................................. 3
2. Considerations .............................................................................................................................. 3
   2.1 Target user group(s) .............................................................................................................. 3
   2.2 Target datasets ...................................................................................................................... 3
   2.3 Data packages vs datasets ................................................................................................... 3
3. Requirements ............................................................................................................................... 3
4. Approach ....................................................................................................................................... 4
5. Submission Workflow .................................................................................................................. 5
6. Metadata Elements ...................................................................................................................... 9
1 Introduction
This document sets out the approach followed during the design of the Data Submission service as part of the EMODnet Data Ingestion portal and it documents the set of metadata elements that has been selected for the data submission form.

2 Considerations

2.1 Target user group(s)
The primary focus of the EMODnet Data Submission service is on data providers and their datasets that are not yet handled and part of the mainstream processes of the data repositories (NODCs, Hydrographic Offices, Geological Services, Biological institutes, etc) that are maintaining the European infrastructures (SeaDataNet, EurOBIS, EGDI, ICES, COGEA) and from there driving the EMODnet Thematic portals (Chemistry, Geology, Bathymetry, Biology, Physical Oceanography and Human Activities).

Potential data providers are marine data holders that are not yet routinely submitting data sets to national data repositories and they must be encouraged and supported to submit their data packages for open access and use in national repositories and EMODnet. It implicates that these providers are not (yet) used to practices and standards as used by the international marine data management community.

2.2 Target datasets
The focus is on datasets that are relevant for the EMODnet Thematic portals. In any case the focus is on datasets including associated documentation, if available, and not on stand-alone reports. These datasets should concern observations and/or analysed samples for chemistry/ geology/ bathymetry/ biology/ physics/ human activities.

2.3 Data packages vs datasets
Identified data providers will be encouraged to submit data packages which rather are collections of datasets, for example resulting from a project, cruise, monitoring programme etc., and not individual datasets.

3 Requirements
A key to success is to identify the right balance between metadata quality, completeness and the effort required by all stakeholders to gather and publish all the information needed. The following requirements were identified:

1. Minimize the effort required from the data provider
It is required that the threshold for submission must be relatively low in order to be successful, especially taking into account that data providers will act on a voluntary basis. It should be made easy for potential data providers to understand and to provide input for the data submission metadata. A commonly agreed principle is that a lot of metadata elements needed, might lead to user frustration and low metadata quality.

2. Utilize data centres’ know-how
Data centres are responsible for data curation and carry a lot of experience and expertise. Data curation involves and implies examination and evaluation of the datasets submitted. In this context, metadata elements, otherwise difficult for the target user group to identify, are easier to be documented by the data centre, since it will be actually part of the data curation procedure.

3. Quality and Automation
It can actually take significant human efforts and time for (post) documenting of incoming data sets with the aim to make these part of data repositories and subsequent sharing with the EMODnet Data portals.
Therefore, metadata elements which can be automatically generated by the system or than can be offered with default values or in pre-completed form were identified.

4. Structured Information

Adoption of controlled vocabularies and limitation of the free text fields to the maximum extent possible is required in order to capture structured information at its source. For that purpose the controlled vocabularies and European directories, both developed and maintained as part of SeaDataNet (www.seadatanet.org) have been adopted.

5. Standards compliancy

A modern system should take into account and benefit from standardization practices. In this context, the ISO 19115-2 standards and INSPIRE Metadata Regulation where adopted as basis for the submission metadatabase, also including elements for the tracking service.

6. Existing good practices

Other ingestion services such as NCEI S2N (https://www.nodc.noaa.gov/submit/) and the Marine Data Exchange of Crowne Estate (UK) (http://www.marinedataexchange.co.uk/) provide good examples and both also make use of the ISO 19115-2 schema.

The Crowne Estate ingestion system conforms to the MEDIN discovery metadata standard v2.3.7 which to a large extent makes use of SeaDataNet Controlled Vocabularies.

The NCEI S2N ingestion makes use of NOAA Controlled Vocabularies and lists (e.g. for organisations, roles, projects, ship/platform, sea area/region, parameter/variable, units, observation category, sampling instrument). The NCEI S2N form gives a nice solution how to make use of controlled lists in combination with free entries and moreover how to enter and save multiple entries.

However, the context for both example services is somewhat different than in the case of EMODnet Ingestion. NCEI S2N is set up for researchers that contractually have to report and submit their data sets to NCEI, while the Marine Data Exchange is set up for consultants and researchers that also somehow are obliged by contract to submit their data sets to Crowne Estate. This implicates that both NCEI and Crowne Estate can set higher requirements for completing the metadata submission forms as well as for the included data packages.

4 Approach

Taking all requirements and considerations into account, the following approach has been adopted for the EMODnet Data Submission service:

1. Make use of ISO 19115-2 and supported by SeaDataNet controlled vocabularies, EDMO, CSR and EDMERP where possible and relevant;
2. Adopt the approach of NCEI S2N for choosing from the controlled lists, adding new terms and entering multiple terms;
3. Make a set-up that the submission form is completed in two steps:
   a. data submitters are asked to complete and submit a subset of the form, including uploading the actual data package consisting of datasets and documentation
   b. the data centre put in charge for further processing will analyze the submission and will get into contact with the data provider for further details. Thereafter the data repository in charge must be able to complete the submission form, using the further insights gathered, and where possible, to work up the received data sets.
   c. The terms used in the submission form must be user friendly and not the technical terms as used in ISO and INSPIRE documentation. Moreover, simple but effective help texts must be added so that submitters can understand easily what to do.
d. Data submitters should learn not to upload too large data collections per data package because of upload performances, but also to restrict having too many different data sets in a data package. Submitters are advised to divide data packages at least by theme and possibly by observation methods in order to get smaller and more harmonised data packages. This implicates that submitters can build series of data packages for specific projects which will be beneficial for the further processing and uptake in data repositories.

5 Submission Workflow
The need for a simple submission mechanism with a rather straightforward online form for specifying relevant metadata to describe the data package, is identified as a success factor for the adoption of the system services to be developed. A typical workflow consists of continuous steps, as described below:

Step1 - Authentication: The data submitter MUST be authenticated via the MarineID service in order to be able to use the Submission service. In this context, the data submitter will be asked if he/she already owns a valid MarineID. In case he/she owns one, he/she will be prompted to enter his/her credentials, gaining access to the Submission service. In case he/she does not own a valid MarineID, he/she will be redirected accordingly to obtain one. The process is automated and the data submitter will get his/her credentials immediately 24/24 and 7/7. The usage of MarineID is very important, in order to have “globally” identified users and server interoperability between the marine systems.

Step 2 - Authorization: After successfully authenticating into the system, user permissions will be checked in order to check the allowed and/or mandatory action the data submitter may perform against the system. Note: the MarineID system is also used for giving access to the EMODnet data centres that will be in charge of processing submitted data packages and completing the submission forms.

Step 3 - Metadata entry: A metadata form will appear for data submitters selecting new submission or edit previous submissions1. The data submitter will have its account to oversee all his/her data submission forms with their status. The metadata form can be seen as an envelope around the actual data package, allowing the documentation of the data package. The selected metadata elements are described in detail in Section 6. Three (3) categories are distincted, namely: (a) mandatory/conditional, (b) optional, (c) system generated.

Metadata elements are grouped in conceptually similar units, organized in tabs/sections etc. Each metadata unit contains both mandatory/conditional and optional elements which fall under the same conceptual category. Visual highlights allow the distinction between mandatory and optional fields. Validation against the corresponding rules established per field is performed. The User Interface has been designed with usability, providing hints about the validation rules and useful error messages for every field completed by the user.

The data submitter will only have to complete a subset of the submission form, part 1, while the assigned data centre will later complete the form part 2, probably in communication with the data submitter.

A general validation against all the rules established will also take place upon the form finalization. Temporary saving of the information entered will be allowed, allowing the user to return at a later stage for completing information missing etc. Furthermore, the “back-end” of the form submission mechanism will check the submitted information, once more, to protect the system from being compromised (sql injection etc.)

Step 4 – Data submission: After successfully filling the required metadata fields, the data submitter will be able to upload data packages as compressed zip file.

---

1 Re-submission is not allowed. Data submitters are able to perform a temporary save, in order to complete or review the submission form at a later stage. When sure, data submitters will perform a final submission and, after an automatic validation check, data packages and forms will be stored with another status, ready for further processing by the data repository that will be put in charge per specific data package.
Finalizing the data submitter part of the submission form will not take place:

- until the corresponding data package is also uploaded
- unless user agrees with open data policy mandated by the project guidelines

**Step 5 – Confirmation:** Data packages will be uploaded at the EMODnet Data Ingestion cloud infrastructure, where it will be stored as secure digital archive. To safeguard security of the ingestion system, proper automated anti-virus and when necessary, manual review workflow has been implemented. The manual review might avoid improper or malevolent data submission not detectable otherwise. A unique ticket number (UUID) for each submission is generated by the system allowing further monitoring and managing of the submission.

**Following steps by data centre in communication with the data submitter:**

**Step 6 – Check by EMODnet Ingestion portal managers:** The portal managers will be alerted by email of the new submission and will check its content to decide which data centre is the best suited for further processing. This decision is made using the country origin of the data submitter and data discipline. In case of multiple disciplines a leading data centre will be selected. The portal managers also might decide to reject the submission which will be underpinned with a motivation. In that case the data submitter is alerted by email and is invited to amend the submission form and related dataset package for a new confirmation (see step 5). This loop can be repeated. Once all is ok then the data centre assignment process of step 6 is continued.

**Step 7 – Assignment of data centre:** The received data submission form and data package will be routed to a data centre, assigned by the EMODnet Ingestion portal managers. An email alert will be send to all contactpersons that are registered in the Submission service for the given data centre. One of the contacts then will assign the specific data centre contactperson that will manage the further processing. The assigned data centre contactperson will be in charge for analysing and processing the data package. This will happen in 2 phases:

- Phase I: from data submission to publishing of the submitted datasets package ‘as is’;
- Phase II: further elaboration of the datasets package and integration (of subsets) in national, European and EMODnet thematic portals.

**Step 8 – Phase I processing by data centre:** The data centre contactperson will download and check the datasets package (datasets collection and associated documentation) and use this to review the submission form part 1 and to complete the submission form part 2. The data centre will contact the data submitter in case of questions. Information about processing actions will be maintained by the data centre in a phase I log in the submission form, so that all is logged and that the data submitter can follow it as well. The data centre will also undertake efforts to replace free texts (‘orphans’), entered by the data submitter, with controlled terms from the SeaDataNet vocabularies and directories. In case these are not yet available, data centre will propose new entries as appropriate. The data centre also might decide to reject the submission which will be underpinned with a motivation. In that case the data submitter is alerted by email and is invited to amend the submission form and related dataset package for a new confirmation (see step 5). This loop can be repeated. Once all is ok then the processing by the data centre as part of step 8 is continued.

**Step 9 – Phase I requesting approval for publishing:** The data centre will validate the submission form checking completeness of obligatory fields for both part 1 and part 2 and will follow-up any issues. The data centre will then request approval of the data submitter for publishing the complete submission form (part 1 and 2) and the original data package. The data submitter will be alerted by email and has 2 weeks for response; otherwise approval is given automatically. The data submitter can request the data centre for an amendment which is entered in the submission log for the data centre to undertake action. The data centre is alerted by email, can

---

2Open Data Policy is adopted. E.g. licensing schemas, commonly used Open Data licenses (Creative commons, Open Data Commons, Directive 96/9/EC on the ‘legal protection of databases’ (Eurlex))
apply the amendment in the submission form and resubmit the request for approval of publishing. This loop can be repeated. Both the data submitter and data centre can retrieve the original part 1 in case of discussion. Once the agreement is there, the submission form and dataset package are released for publishing.

**Step 10 – Phase I final check for publishing:** The portal managers will be alerted by email of the release for publishing and will undertake a final review. If ok, a confirmation will be entered, that will trigger the exchange of submission form information by means of XML from the Ingestion Submission service to the Ingestion Discovery & Access Service.

**Step 11 – Phase I publishing and completion:** After successful ingestion of the XML information into the Discovery and Access service, then the submission form and associated datasets package (datasets collection and documentation) ‘as is’ will be publicly available for searching and downloading in the online Discovery & Access Service. Once the submission is public, the portal managers will complete the tracking info in the Submission service with an optional DOI (via DataCite) in case of academic publishing and the publication link.

**Following steps by data centre in communication with the data submitter for completing Phase II:**

Phase II deals with further elaboration by the assigned data centre of the datasets package and integration (of subsets) in national, European and EMODnet thematic portals. This will be done possibly in further contact with Data Submitter.

**Step 12 – Phase II data centre processing of datasets:** The assigned data centre will further process and curate the submission at dataset level and will work up (subsets of) the received datasets to the level of inclusion in the data management system of the data centre, including detailed metadata. For this purpose the data centre might have extra communication with the data submitter and will maintain also an extra log in the Submission service, documenting processing steps. The log will again be accessible for data submitter, data centre, and Ingestion portal managers from the start of processing. Finalised datasets (possibly parts of original data package) will be included in the data management system of the data centre and also published at the data centre portals and taken up for long term stewardship.

**Step 13 – Phase II European and EMODnet publishing:** The data centre will populate the finalised curated datasets and related metadata into the appropriate European infrastructures (SeaDataNet, EuroBIS, EGDI, ICES, COGEA). The finalised datasets and related metadata will be pushed forward from the European infrastructures towards inclusion and publishing in EMODnet thematic portals. Once done, the data centre will complement the data submission form with details of the national, European and EMODnet URLs. The Ingestion portal managers will be alerted and include these links also in the Public Discovery and Access service.

Note: Phase II is under development in the Submission service.
Figure 1: Simplified Submission Workflow for the part 1 of the submission form that is completed by the data submitter
### 6 Metadata Elements

The metadata table detailed below is the result of an intensive revision process between project partners.

1. Metadata elements are grouped in conceptual units
2. Some metadata elements are ‘merged’ in the table in order to appear similar to the way they will appear in the submission form
3. Metadata elements are split over actors based on how they are completed:
   a. Data Submitter (DS), these elements together are part 1 of the submission form (excl AUTO and AUTO Default elements)
   b. Data Centre (DC), these elements together are part 2 of the submission form (excl AUTO and AUTO Default elements)
   c. Automatic (AUTO),
   d. Default values (AUTO Default)
4. The naming schema is distinguished on:
   a. ISO/INSPIRE used for database
   b. Public naming as seen on Submission Form
5. Some metadata elements were added in order to facilitate the publication workflow:
   a. Date of metadata latest revision
   b. URL where the dataset can be downloaded
   c. Instrument type
   d. Instrument name
   e. Publication Url
   f. Publication Date
   g. DOI
6. Missing are some extra metadata elements which will be added because of Phase II which is under development.

<table>
<thead>
<tr>
<th>Metadata Element</th>
<th>ISO/TS 19139 path</th>
<th>Metadata Element Submission Form</th>
<th>Description</th>
<th>Hint</th>
<th>Filed in by</th>
<th>Mandatory</th>
<th>Multiplicity</th>
<th>Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metadata point of contact</td>
<td>contact</td>
<td>Contact person for this dataset submission form</td>
<td>Organization/person responsible for the creation and maintenance of the metadata. This refers to the metadata record created during submission and not to the responsible party for the resource, since they will not necessarily be the same. It will be automatically filled, using the user’s credentials used for authentication.</td>
<td>Information shown is based on your MarineID profile. In case some information is outdated and/or incomplete, please use the “Edit Profile” function on the top right of the screen</td>
<td>DS</td>
<td>M</td>
<td>1</td>
<td>AUTO Default: English</td>
</tr>
<tr>
<td>Responsible party and role</td>
<td>identificationInfo[0]/pointOfContact</td>
<td>Organizations responsible for the dataset</td>
<td>Description of the organization responsible for the establishment, management, maintenance or distribution of the resource. Provide identification of, and means of communication with, person(s) and organization(s) associated with the resource(s). Rule of the responsible party. For each role related to the provided, an organization should be also be provided. The user will be prompted to use the most applicable role.</td>
<td>Provide the organisations that are responsible for the collection and the management of the dataset. Fill in one organisation &amp; corresponding role at the time.</td>
<td>DS</td>
<td>M</td>
<td>1</td>
<td>Country: ISO 3166; Organization: EDMO or free text</td>
</tr>
<tr>
<td>Metadata language</td>
<td>language</td>
<td>Language used for completing form</td>
<td>Language used for completing form</td>
<td>Is shown on the DS submission screen so that the DS is made aware.</td>
<td>AUTO</td>
<td>M</td>
<td>1</td>
<td>CodeList (see ISO/TS 19139) based on alpha-3 codes of ISO 639-2</td>
</tr>
<tr>
<td>Resource title</td>
<td>identificationInfo[0]/citation/*/title</td>
<td>Title of dataset</td>
<td>A characteristic, and often unique, name by which the resource is known. The Resource Title has to be concise and to the point. It should not contain unexplained acronyms or abbreviations. It is recommended a maximum length of 250 characters</td>
<td>Enter a concise title which characterises well the subject of the dataset. A maximum of length of 250 characters is recommended</td>
<td>DS</td>
<td>M</td>
<td>1</td>
<td>Free text</td>
</tr>
</tbody>
</table>
| Resource abstract | identificationInfo/*/abstract | Narrative summary of dataset | A brief narrative summary of the content of the resource. The abstract provides a clear and concise statement that enables the reader to understand the content of the data.

The resource abstract is a succinct description that can include:
- A brief summary with the most important details that summarise the dataset submitted
- Objectives of data collection
- Coverage: In situ in the extent or location in addition to the bounding box
- Main attributes
- Data sources
- Legal references
- Importance of the work

It is advised to summarise the most important details in the first sentence or first 100 characters.
Encourage submitters to provide a meaningful summary.

Provide a summary of the content of the data package. The summary may include:
- The most important details that summarise the dataset submitted
- Objectives of data collection
- Area of data collection
- Main attributes
- Data sources
- Legal references
- Importance of the work

Encourage submitters to provide a meaningful summary.

| Resource format | /distributionInfo/* | Dataset format | Listing of (suggested) common and interoperable formats

Use mechanism of S2N to search and include 124 terms or free text

Choose one or more formats and/or enter new one(s) if the applicable format is not listed.

| Resource language | identificationInfo/*/language | Language used in dataset | ISO 19115 mandates the dataset language, even if the resource does not include any textual information. The ISO 19115. Dataset language is defaulted to the Metadata language (see below)

Element not shown on Submission Form

Shown on data centre and publication form

Default: English

| Lineage | dataQualityInfo/*/lineage/*/statement | Summary of data quality processing | General explanation of the data producer’s knowledge about the lineage of a dataset. A statement on process history and/or overall quality of the spatial data set. Where appropriate it may include a statement whether the data set has been validated or quality assured, whether it is the official version (if multiple versions exist), and whether it has legal validity.

Mandatory for INSPIRE

Provide information regarding the processing history and an overall quality statement of the dataset. In the latter case, please be sure to provide accompanying documentation below.

| Documentation | /distributionInfo/* | Relevant supporting documentation -Document | Supporting documentation:
In cases that no common (suggested) data formats are adopted, data providers will be asked to include also documentation about their formats and coding in the data packages

Data submitter must be informed well what might be included and must be requested to include the documents in the Data Package that will be submitted

Fill in the supporting documentation and do not forget to include the documents in the submission package. The documents will be very helpful for the further handling of your dataset by the data repository in charge.

| Limitations on public access | /resourceConstraints/*/accessConstraints | Public access | Information on the limitations and the reasons for them.

If there are no limitations on public access, use the free text available in other Constraints to enter “No Limitations” in the language used for the metadata.

For EMODnet Ingestion this is a prerequisite.

All datasets submitted to the EMODnet Ingestion portal will be made publicly available

| Conditions applying to access and use | /resourceConstraints/*/useLimitation | License for use | Conditions for access and use

Choose an Open License or enter a License which should be compatible with open data definition

| Conditions applying to access and use | /resourceConstraints/*/useLimitation | License for use | Conditions for access and use

Choose an Open License or enter a License which should be compatible with open data definition

| Attribution License (ODC-BY) | - Attribution License (ODC-BY)

- CC0 1.0 Universal – public domain dedication

- Creative Commons Attribution 4.0 International (CC-BY 4.0)
<table>
<thead>
<tr>
<th>Resource Type</th>
<th>hierarchyLevel</th>
<th>Type</th>
<th>Description</th>
<th>Metadata date</th>
<th>Resource code</th>
<th>Resource locator</th>
<th>Data Types</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cruise</td>
<td></td>
<td></td>
<td>Classification of the resource based on its scope  - dataset is an identifiable data that can be accessed separately. A dataset can be a part of a whole (series) or a separate resource  - series is a collection of resources or related datasets that share the same product specification</td>
<td>The submitted data package should comprise an identifiable dataset</td>
<td>AUTO</td>
<td>Default: Dataset</td>
<td>M</td>
</tr>
<tr>
<td>Resource locator</td>
<td>distributionInfo/*</td>
<td>citation/*</td>
<td>Points users to the location (URL) where the data can be downloaded, or to where additional information about the resource may be provided</td>
<td>URL pointing at submission service cloud storage</td>
<td>DC</td>
<td>M</td>
<td>1</td>
</tr>
<tr>
<td>Resource code</td>
<td>identificationInfo/*</td>
<td>citation/*</td>
<td>Value uniquely identifying the resource. Code used for external references, product code for users</td>
<td>Element not shown on Submission Form for DS</td>
<td>Show on data centre and publication form</td>
<td>DC</td>
<td>AUTO</td>
</tr>
<tr>
<td>Metadata date</td>
<td>dateStamp</td>
<td>Date of metadata creation</td>
<td>The date which specifies when the metadata record was created or updated</td>
<td>The field is automatically populated with the date of submission</td>
<td>AUTO</td>
<td>Default: date of final form submission</td>
<td>M</td>
</tr>
<tr>
<td>Metadata date</td>
<td>dateStamp</td>
<td>Date of metadata latest revision</td>
<td>The date which specifies when the metadata record was updated by the data centre</td>
<td>The field is automatically populated with the date of last revision by the data centre</td>
<td>AUTO</td>
<td>Default: date of form edit by DC</td>
<td>M</td>
</tr>
</tbody>
</table>

### Data Types

- **M**o**ッション**
  - **Project / Programme**
    - Identifiable activity which provided the data
    - Use mechanism of S2N to search and include EDMERP title (and code)
    - Inform data submitter to submit one or more Data Packages for data from one project or programme and not for multiple projects in one go
    - Please provide the name of the project / programme in which the dataset was collected
    - Choose one title from the list or enter a new one if the applicable project/programme is not listed.
    - Element not shown on Submission Form
    - Show on data centre and publication form
    - Hint for DC: Choose one or more Cruise names from the CSR list and/or enter new ones.
    - Default: EDMERP (European Directory of Marine Environmental Research Projects)
    - **OR**
    - Free text
  - **Cruise**
    - **Cruise Summary Report**
      - Cruise details
      - Use mechanism of S2N to search and include CSR titles (and codes)
      - Element not shown on Submission Form
      - Show on data centre and publication form
      - Hint for DC: Choose one or more Cruise names from the CSR list and/or enter new ones.
      - Default: Cruise Summary Report Inventory (OR)
      - **OR**
      - Free text
  - **Process Methodology**
    - **Summary of processing methodology**
      - Comprehensive information about the procedure by which the algorithm was applied to derive geographic data from the raw instrument measurements, such as datasets, software used, and the processing environment
      - Provide information about the procedure used to derive observation data from the instrument measurements, including possible software used, and applied processing steps
      - Default: Free text
  - **Platform Type**
    - **Platform**
      - Designation of the platform type used to acquire the dataset
      - Use mechanism of S2N to search and include L06 names (and codes)
      - Element not shown on Submission Form
      - Show on data centre and publication form
      - Hint for DC: Choose one or more platform types from the list; complete type and name in combinations
      - Default: SEAVOX Platform Categories (OR)
      - **OR**
      - ICES Platform Codes
  - **Platform Name**
    - **Platform name**
      - Designation of the platform used to acquire the dataset
      - Element not shown on Submission Form
      - Default: Free text

---

- **Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)**
- **Open Database License (ODC-ODDL)**
- **Public Domain Dedication and License (PDDL)**
<table>
<thead>
<tr>
<th>Parameter/Variable</th>
<th>Description</th>
<th>Choice</th>
<th>DC</th>
<th>O</th>
<th>1.*</th>
<th>1..*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Observation type</td>
<td>Ocean parameters processed in the current dataset. Only applicable for phenomenons values, not applicable for coordinates (e.g. lat, lon), for values metadata e.g. errors, quality flags) or technical parameters (e.g. analysis residual, battery charge, ...) Use mechanism of S2N to search and include P03 terms (and codes)</td>
<td>Choose the applicable parameter groups from the list</td>
<td>DS</td>
<td>M</td>
<td>1..*</td>
<td>1..*</td>
</tr>
<tr>
<td>Parameter/Variable</td>
<td>The physical, biological, geophysical, or chemical property being measured or calculated, e.g., water temperature. Could also relate to human activities For each variable, units, category and instrument shall be provided Use mechanism of S2N to search and include L22 terms (and codes)</td>
<td>Choose the applicable discovery parameters from the list</td>
<td>DC</td>
<td>M</td>
<td>1..*</td>
<td>1..*</td>
</tr>
<tr>
<td>Topic category</td>
<td>A high-level classification scheme to assist in the grouping and topic based search of available spatial data resources for compliance with INSPIRE and defaults to &quot;Oceans&quot; for datasets falling under &quot;Oceanographic geographical features&quot;</td>
<td>Element not shown on Submission Form, data centre and publication form Only in XML and database</td>
<td>AUTO: Default: Oceanographic geographical features</td>
<td>M</td>
<td>1..*</td>
<td>1..*</td>
</tr>
<tr>
<td>Keyword Value - INSPIRE</td>
<td>INSPIRE Data theme. Defaults to &quot;Oceanographic geographical features&quot;</td>
<td>Element not shown on Submission Form, data centre and publication form Only in XML and database</td>
<td>AUTO: Default: Oceanographic geographical features</td>
<td>M</td>
<td>1..*</td>
<td>1..*</td>
</tr>
<tr>
<td>Originating controlled vocabulary - INSPIRE</td>
<td>Name of the formally registered thesaurus. Defaults to “GEMET - INSPIRE themes, version 1.0, publication, 2008-06-01” for INSPIRE data themes</td>
<td>Element not shown on Submission Form, data centre and publication form Only in XML and database</td>
<td>AUTO: Default: GEMET - INSPIRE themes, version 1.0, publication, 2008-06-01</td>
<td>M</td>
<td>1..*</td>
<td>1..*</td>
</tr>
<tr>
<td>Locations &amp; Dates</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Geographic Bounding Box</strong></td>
<td>Extent of the resource in the geographic space, given as a bounding box. The coordinates of the bounding box are expressed in any geodetic coordinate reference system with a Greenwich Prime Meridian in decimal degrees with at least 3 decimal digits.</td>
<td><strong>DS</strong></td>
<td><strong>M</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spatial resolution</strong></td>
<td>Spatial resolution refers to the level of detail of the dataset. It shall be expressed as a set of zero to many resolution distances (typically for gridded data and imagery-derived products) accompanied with units of measure.</td>
<td><strong>DC</strong></td>
<td><strong>O</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Vertical extent of dataset</strong></td>
<td>Vertical extent contained in the dataset. It will be provided as highest and lowest values with units of measurement.</td>
<td><strong>DC</strong></td>
<td><strong>O</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Temporal extent</strong></td>
<td>Time period covered by the content of the dataset. It will be provided as start and end dates.</td>
<td><strong>DS</strong></td>
<td><strong>M</strong></td>
<td><strong>1</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coordinate Reference System**

- **Reference System**
  - **Reference System as used in the dataset**
  - **Coordinate Reference System used in dataset**

**Sea areas keyword**

- **New sea area**
  - **Sea area name**
  - **Sea areas covered by the dataset**

**Sea areas originating controlled vocabulary**

- **New sea area**
  - **Sea area name**
  - **Sea areas covered by the dataset**

**Geographic Bounding Box**

- **North**
- **South**
- **East**
- **West**

---

**Temporal extent**

- **Start date**
- **End date**

**Spatial resolution**

- **Spatial resolution unit**
- **Spatial resolution value**

**Vertical extent**

- **Minimum value**
- **Maximum value**
- **Unit of Measure**

---

**Coordinate Reference System**

- **Coordinate Reference System used in dataset**
- **Coordinate Reference System selected**
- **Coordinate Reference System selected**

**Sea areas keyword**

- **New sea area**
- **Sea area name**
- **Sea areas covered by the dataset**

---

**Vertical domain of dataset**

- **Minimum value**
- **Maximum value**
- **Unit of Measure**

---

**Spatial resolution**

- **Spatial resolution unit**
- **Spatial resolution value**
- **Unit of Measure**

---

**Temporal extent**

- **Start date**
- **End date**
<table>
<thead>
<tr>
<th>Specification</th>
<th>INSPIRE Conformance</th>
<th>Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>dataQualityInfo/report/<em>result/</em></td>
<td>INSPIRE mandated element. Can be automatically generated as the submitter is supposed to have knowledge of INSPIRE specifications.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Three couples specification – degree will be completed in the backend:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Corrigendum to INSPIRE Metadata Regulation published in the Official Journal of the European Union, L 328, page 83</td>
<td>Conformant</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency of observations</td>
<td>Temporal resolution (mean period between each time steps)</td>
<td>DC</td>
</tr>
<tr>
<td></td>
<td>Not mentioned on data submitter form, only on data repository form and later on public pages</td>
<td>O</td>
</tr>
<tr>
<td></td>
<td>Indicate the frequency of time series, choosing from the list</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>DC</td>
<td>03 – SeaDataNet Measurement Periodicity Categories</td>
</tr>
<tr>
<td>Date of publication</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation.</td>
<td>DC</td>
</tr>
<tr>
<td></td>
<td>Since date of creation will be filled in by the data submitter this element will serve to document the publication process of ingestion and safe-keeping of marine data system</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Not mentioned on data submitter form, only on data centre form and later on public pages</td>
<td>1</td>
</tr>
<tr>
<td>Date of last revision</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation.</td>
<td>DC</td>
</tr>
<tr>
<td></td>
<td>Element not shown on Submission Form</td>
<td>C</td>
</tr>
<tr>
<td>Date of creation</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation</td>
<td>DS</td>
</tr>
<tr>
<td></td>
<td>Indicate the date the package was created</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Date that the original dataset and the completed data submission form are published and available to the public through the EMODnet Ingestion portal</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Do not expose on public form.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Not mentioned</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day)</td>
<td></td>
</tr>
<tr>
<td>Date of publication</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation.</td>
<td>DC</td>
</tr>
<tr>
<td>Date of last revision</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation.</td>
<td>DC</td>
</tr>
<tr>
<td>Date of creation</td>
<td>To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation</td>
<td>DC</td>
</tr>
<tr>
<td></td>
<td>Indicate the date the package was created</td>
<td>M</td>
</tr>
<tr>
<td></td>
<td>Date that the original dataset and the completed data submission form are published and available to the public through the EMODnet Ingestion portal</td>
<td>1</td>
</tr>
</tbody>
</table>

**Temporal resolution**

<table>
<thead>
<tr>
<th>Date of creation</th>
<th>Date of last revision</th>
<th>Date of publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date of publication</td>
<td>IdentificationInfo/md:DOI/md:emento/md:object/md:creator/md:core/md:creation</td>
<td>Date of creation</td>
</tr>
<tr>
<td>Date of last revision</td>
<td>IdentificationInfo/md:DOI/md:emento/md:object/md:creator/md:revision</td>
<td>Date of last revision</td>
</tr>
<tr>
<td>Date of creation</td>
<td>IdentificationInfo/md:DOI/md:emento/md:object/md:creator/md:creation</td>
<td>Date of creation</td>
</tr>
</tbody>
</table>

**INSPIRE Conformance**

- Specification
- Data Quality Info
- Report
- Result
- M: Conformant
- O: Not conformant
- L*: Conditional, from perspective of processing or not in data centre
- M: Conditional, from perspective of processing or not in data centre
- L*: Conditional, from perspective of processing or not in data centre
## Upload & Finish

<table>
<thead>
<tr>
<th>Uploaded file name</th>
<th>Ingestion Service specific</th>
<th>Filename</th>
<th>Filename of the data package submitted</th>
<th>Indicate the filename of the data package submitted with this form</th>
<th>DS</th>
<th>M</th>
<th>1</th>
<th>Free text</th>
</tr>
</thead>
</table>

## Process info

<table>
<thead>
<tr>
<th>UUID</th>
<th>Ingestion Service specific</th>
<th>Submission identifier (UUID)</th>
<th>Submission identifier automatically assigned by the Ingestion Service</th>
<th>DS</th>
<th>AUTO</th>
<th>RFC 4122, ISO/IEC 9834-8:2005</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Submission Status</th>
<th>Ingestion Service specific</th>
<th>Status</th>
<th>Status indicating the process stage of a submission</th>
<th>DS, DC, M</th>
<th>M</th>
<th>1</th>
<th>10 = Drafting form part 1 by Data Submitter 15 = Form rejected by Master and returned to Data Submitter 20 = Form part 1 submitted by Data Submitter 25 = Leading Data Centre assigned by Master 30 = Data Centre Contact assigned by Data Centre 40 = Completing parts 1 and 2 by Data Centre Contact 50 = Approval for publishing requested from Data Submitter 60 = Amendment requested by Data Submitter 70 = Submitted for publishing by Data Centre 75 = Publication draft 80 = Sent to Discovery and Access service 85 = Published at Discovery and Access service 90 = Rejected by Master and returned to Data Submitter 91 = Rejected by Data Centre and returned to Data Submitter</th>
</tr>
</thead>
</table>

| Leading Data Centre | Ingestion Service specific | Assigned Data Centre | Leading Data Centre (assigned by Master) | - | M | M | 1 | Controlled list of Data Centres registered at the Ingestion Service |

| Data Centre Contact | Ingestion Service specific | Contact Data Centre | Data Centre point of contact responsible for reviewing the form | - | DC | M | 1 | Controlled list of users of the Assigned Centre |

| Form Last update | Ingestion Service specific | Last Update | Timestamp of latest form revision | AUTO | M | 1 |

| Resource locator | distributionInfo/*/transferOptions/*/onLine/*/linkage | Dataset URL | Points users to the location (URL) where the data can be downloaded, or to where additional information about the resource may be provided | Shown on data centre and publication form | AUTO; Link to Ingestion Service Cloud Storage | M | 1 | URL (IETF RFC2738 and IETF RFC 2056) |

| Date of publication | identificationInfo[1]/citation/*/data| Date of dataset publication | To be compliant with ISO 19115 it is necessary to use at least one among date of publication, date of last revision, or the date of creation. | Date that the original dataset and the completed data submission form are published and available to the public through the Discovery and Access Service | M | M | 1 | The default reference system shall be the Gregorian calendar, with dates expressed in accordance with ISO 8601 (yyyy-mm-dd where yyyy is the year, mm is the month and dd is the day) |

| Publication URL | distributionInfo/*/transferOptions/*/onLine/*/linkage | Publication URL | Url of the submission | Date that the original dataset and the completed data submission form are published and available to the public through the Discovery and Access Service | M | M | 1 | URL (IETF RFC2738 and IETF RFC 2056) |

| DOI | identificationInfo[1]/citation/*/id | DOI | DOI in case of academic publication | Identifier of the publication | M | O | 1 |