EMODnet Data Ingestion and safe-keeping of marine data

Submission Form - Workflow procedure and Metadata Elements

Version: V2.0.0

Date Original: 5 February 2017
Latest Update: 13 March 2025

Prepared by: HCMR, MARIS and GET (subcontractor of HCMR) with input of consortium partners

Table of Contents

1 Introduction

This document sets out the approach followed during the design of the Data Submission service as part of EMODnet Data Ingestion 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 providing data input to the EMODnet Thematic groups (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 groups. 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 volontary 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 EMODnet.

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 were 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 analyse 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 ECAS AAI 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 ECAS login. 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 ECAS login, 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.

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 ECAS AAI 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 submissions[1]. 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.

Finalizing the data submitter part of the submission form will not take place:

- until the corresponding data package is also uploaded

[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.

- unless user agrees with open data policy mandated by the project guidelines[2]

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 managers: The 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 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 managers. An email alert will be send to all contact persons that are registered in the Submission service for the given data centre. One of the contacts then will assign the specific data centre contact person that will manage the further processing. The assigned data centre contact person 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, and European infrastructures which feed data to EMODnet thematic groups and services.

Step 8 – Phase I processing by data centre: The data centre contact person 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 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.

[2] Open 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))

Step 10 – Phase I final check for publishing: The 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 EMODnet Products Catalogue Service.

Step 11 – Phase I publishing and completion: After successful ingestion of the XML information into the EMODnet Products Catalogue 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 EMODnet Products Catalogue Service. Once the submission is public, the managers will complete the tracking info in the Submission service with 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 and European infrastructures which data to the EMODnet thematic groups and services. 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 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 finalized datasets and related metadata will be pushed forward from the European infrastructures towards the EMODnet thematic groups. Once done, the data centre will complement the data submission form with details of the national, European and EMODnet URLs. The Ingestion managers will be alerted and include these links also by updating the EMODnet Products Catalogue Service entries.

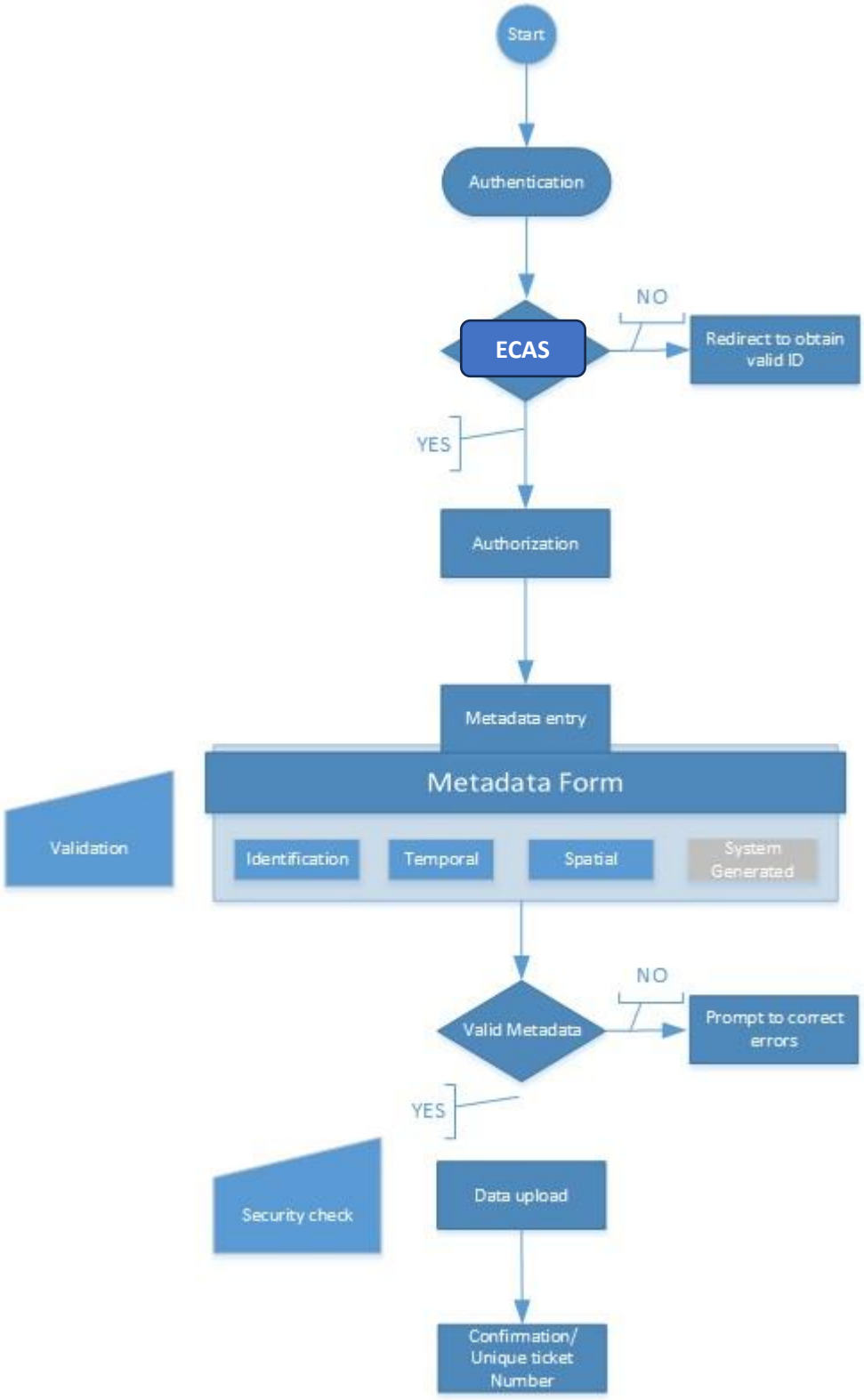

Figure 1: Simplified Submission Workflow for the part 1 of the submission form that is completed by the data submitter

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 concerning Phase II.

Metadata Element ISO/INSPIRE	ISO/TS 19139 path	Metadata Element Submission Form	Description	Hint	Filled in by	Mandatory	Multiplicity	Domain
Organizations								
Metadata point of contact	contact	Contact person for this dataset submission form • Name • Surname • Email • Country • Organization	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.	Information shown is based on your ECAS Login profile. In case some information is outdated and/or incomplete, please use the “Edit Profile” function on the top right of the screen	DS	M	1	• Name: Free text • Surname: Free text • Email: Free text • Country: ISO 3166 • Organization: EDMO or free text
Responsible party and Role	identificationInfo[1]/*/pointOfContact	Organizations responsible for the dataset • Country • Organization • Email of organization • Role of organization	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) Role 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	Provide the organisations that are responsible for the collection and the management of the dataset. Fill in one organisation & corresponding role at the time.	DS	M	1..*	• Country: ISO 3166 • Organization: EDMO or free text Email of organization: free text Role of organization: - Dataset Holding Organisation - Originator of Dataset
Metadata language	language	Language used for completing form	Language used for completing form Is shown on the DS submission screen so that the DS is made aware		AUTO Default: English	M	1	Codelist (See ISO/TS 19139) based on alpha-3 codes of ISO 639-2

Dataset Identificatior								
Resource title	identificationInfo[1]/*/citation/*/title	Title of dataset	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	Enter a concise title which characterises well the subject of the dataset. A maximum of length of 250 characters is recommended	DS	M	1	Free text
			and keeping the similarity with the original title of the resource, in the sense of the 'official naming' Encourage submitters to provide a meaningful title					
Resource abstract	identificationInfo[1]/*/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: linguistic transcriptions of the extent or location in addition to the bounding box - Main attributes - Data sources - Legal references - Importance of the work It is advised to summarize 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	DS	M	1	Free text
Resource Format	/distributionInfo/MD_Distribution/*/distributionFormat	Dataset format	Listing of (suggested) common and interoperable formats Use mechanism of S2N to search and include L24 terms or free text	Choose one or more formats and/or enter new one(s) if the applicable format is not listed.	DS	M	1..*	UK MEDIN formats (M01) Or Free text
Resource language	identificationInfo[1]/*/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	DC Default: English	M	1	Default to English, also use English even if language not applicable Codelist (See ISO/TS 19139) based on alpha-3 codes of ISO 639-2
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	DS	M	1	Free text
Documentation	/distributionInfo/MD_Distribution/*/distributionFormat/formatSpecificationCitation/*/	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	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.	DS	O	1..*	Free text

			documents in the Data Package that will be submitted!					
Limitations on public access	identificationInfo[1]/*/resourceConstraints/*/accessConstraints identificationInfo[1]/*/resourceConstraints/*/otherConstraints	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 themetadata For EMODnet Ingestion this is a prerequisite.	All datasets submitted to the EMODnet Ingestion will be made publicly available	AUTO Default: No limitations	M	1	B.5.24 of ISO19115
Conditions applying to access and use	identificationInfo[1]/*/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	DS	M	1	An identifiable collection of open data licenses: -Attribution License (ODC-By) - CC0 1.0 Universal – public domain dedication (CC0 1.0) - Creative Commons Attribution 4.0 International (CCBY 4.0)

								- Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0) -Open Database License (ODC-ODbL) - -Public Domain Dedication and License (PDDL)
Resource Type	hierarchyLevel	Type	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 segregate resource - series: is a collection of resources or related datasets that share the same product specification Shown on screen so that submitters are reminded	The submitted data package should comprise an identifiable dataset	AUTO Default: Dataset	M	1	CodeList (Annex B.5.25 of ISO 19115) Default: Dataset
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	URL pointing at submission service cloud storage	DC	M	1	URL (IETF RFC1738 and IETF RFC 2056) Automatically created by renaming the submitted data package with UUID of the submission
Resource code	identificationInfo[1]/*/citation/*/identifier	Dataset identifiers	Value uniquely identifying the resource. Code used for external references, product code for users Not mentioned on data submitter form, only on data repository form and later on public pages	Element not shown on Submission Form for DS Shown on data centre and publication form	DC AUTO: UUID of submission form	M	1..*	B.2.7.3 of ISO 19115
Metadata date	dateStamp	Date of metadata creation	The date which specifies when the metadata record was created or updated	The field is automatically populated with the date of submission	AUTO Default: date of final form submission	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)
Metadata date	dateStamp	Date of metadata latest revision	The date which specifies when the metadata record was updated by the data centre	The field is automatically populated with the date of last revision by the data centre	DC AUTO: date of form edit by DC	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)
Data Types								

Mission	MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Operation/description	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 Package(s) for data from one project or programme and not for multiple project 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.	DS	O	1	EDMERP (European Directory of Marine Environmental Research Projects) OR Free text
Cruise	MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Operation/*/description MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Operation/*/citation	Cruise Summary Report	Cruise details Use mechanism of S2N to search and include CSR titles (and codes)	Element not shown on Submission Form Shown on data centre and publication form Hint for DC: Choose one or more Cruises from the CSR list and/or enter new ones.	DC	O	1..*	Cruise Summary Report Inventory (CSR) OR Free text
Process Methodology	LI_ProcessStep/*/ LE_ProcessStep/*/ LE_Processing/*/procedureDescription	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	DS	O	1	Free text
Platform Type	MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Operation/*/ MI_Platform/*/ citation	Platform type	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 Shown on data centre and publication form Hint for DC: Choose one or more platform types from the list; complete type and name in combinations	DC	O	1..*	SeaVox Platform Categories (L06)
Platform Name	MI_Metadata/*/ MI_AcquisitionInf	Platform name	Designation of the platform used to acquire the dataset	Element not shown on Submission Form	DC	O	1..*	ICES Platform Codes OR

	ormation/*/MI_Operation/*/ MI_Platform/*/identifier		Use mechanism of S2N to search and include C17platform names (and codes!) Not mentioned on data submitter form, only on data repository form and later on public pages	Shown on data centre and publication form Hint for DC: Choose one or more platform names from the list or enter new ones; complete type and name in combinations				free text
Instrument type	MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Instrument/identifier/type	Instrument type	Use mechanism of S2N to search and include L05 terms (and codes) Complete in combinations of parameters and instrument Not mentioned on data submitter form, only on data repository form and later on public pages	Choose one or more instrument types from the list; complete type and name in combinations	DC	O	1.*	SeaDataNet Device Categories (L05)
Instrument	MI_Metadata/*/ MI_AcquisitionInformation/*/MI_Instrument/identifier	Instrument name	Use mechanism of S2N to search and include L22 terms (and codes)	Choose one or more instrument names from the list; complete type and name in combinations	DC	O	1..*	SeaVoX Device Catalogue

~~Parameter/Variables~~	identificationInfo[1]/*/descriptiveKeywords/*/keyword	Parameter / Variable	The physical, biological, geological, gephysical 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 P02 terms (and codes!)	Choose the applicable discovery parameters from the list	DC	M	1..*	SeaDataNet P02 – Parameter Discovery Vocabulary
Marine organism	identificationInfo[1]/*/descriptiveKeywords/*/keyword	Marine organism	Marine organism	Choose the appropriate taxonomic group by entering free text after looking up in the WoRMS register	DC	O	1..*	Lookup value at WORMS http://www.marinespecies.org/
Ocean discovery parameters keyword	identificationInfo[1]/*/descriptiveKeywords/*/keyword	Observation type	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)	Choose the applicable parameter groups from the list	DS	M	1..*	SeaDataNet Agreed Parameter Groups (P03)
Ocean discovery originating controlled vocabulary	identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName	Ocean discovery originating controlled vocabulary	Name of the formally registered thesaurus. Will default to “SeaDataNet Agreed Parameter Groups, v22, revision, 2015-02-04	Element not shown on Submission Form, data centre and publication form Only in XML and database	AUTO: Default: SeaDataNet Agreed Parameter Groups	M	1	“SeaDataNet SeaDataNet Agreed Parameter Groups, v22, revision, 2015-02-04
Topic category	identificationInfo[1]/*/topicCategory	Topic category	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 “Oceans” for datasets falling under “Oceanographic geographical features” theme	Element not shown on Submission Form, data centre and publication form Only in XML and database	AUTO Default: Oceans	M	1	Enumeration (B.5.27 of ISO 19115)
Keyword Value - INSPIRE	identificationInfo[1]/*/descriptiveKeywords/*/keyword	Keyword Value -INSPIRE	INSPIRE Data theme. Defaults to “Oceanographic geographical features”	Element not shown on Submission Form, data centre and publication form Only in XML and database	AUTO: Default: Oceanographic geographical features	M	1	INSPIRE Data themes GEMET - INSPIRE themes, version 1.0, publication, 200806-01
Originating controlled vocabulary - INSPIRE	identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName	Originating controlled vocabulary - INSPIRE	Name of the formally registered thesaurus. Defaults to “GEMET - INSPIRE themes, version 1.0, publication, 2008-06-01” for INSPIRE data themes	Element not shown on Submission Form, data centre and publication form Only in XML and database	AUTO: Default: GEMET - INSPIRE themes, version 1.0, publication, 200806-01	M	1	Defaults to “GEMET - INSPIRE themes, version 1.0, publication, 2008-06-01” for INSPIRE data themes

Locations & Dates

Geographic bounding box	identificationInfo[1]/*/extent/*/geographicElement/*/westBoundLongitude identificationInfo[1]/*/extent/*/geographicElement/*/eastBoundLongitude identificationInfo[1]/*/extent/*/geographicElement/*/southBoundLatitude identificationInfo[1]/*/extent/*/geographicElement/*/northBoundLatitude	Geographic Bounding Box • North • South West • East	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 PrimeMeridian in decimal degrees with at least 3 decimal digits	Draw a rectangle on the map to indicate the area covered by the dataset	DS	M	1	-180.00 ≤ westBoundLongitude ≤ 180.00 -180.00 ≤ eastBoundLongitude ≤ 180.00 -90.00 ≤ southBoundingLatitude ≤ 90.00 -90.00 ≤ northBoundingLatitude ≤ 90.00
Coordinate reference system	referenceSystemInfo	Coordinate Reference System	Description of the coordinate reference system(s) used in dataset. Highly recommended to consistent across a dataset prior to submission. Show on screen so that submitters are reminded And EPSG:4258 added on the backend	Coordinate reference system used in dataset	AUTO Default: EPSG:4326 (WGS84) EPSG:4258 added on the backend	M	1..* Multiple in XML and database	EPSG codes http://www.opengis.net/def/crs/EPSG/0/4258 Default to EPSG:4326 (WGS84) And EPSG:4258 added on the backend
Sea areas keyword	identificationInfo[1]/*/descriptiveKeywords/*/keyword	Sea area	Sea areas covered by the dataset Use mechanism of S2N to search and include C16 terms (and codes)	Choose sea areas covered by the dataset from the list	DS	O	1..*	SeaVoX salt and fresh water body gazetteer http://vocab.nerc.ac.uk/collection/C19/current/
Sea areas originating controlled vocabulary	identificationInfo[1]/*/descriptiveKeywords/*/thesaurusName	Sea areas originating controlled vocabulary	Name of the formally registered thesaurus. Will default to "SeaVoX salt and fresh water body gazetteer, version 16, revision, 2015-02-19" for sea areas keywords	Element not shown on Submission Form, data centre and publication form Only in XML and database	AUTO: Default: SeaVoX salt and fresh water body gazetteer	O	1	NERC Vocabulary Server version 2.0, revision, 2012-0416 "SeaVoX salt and fresh water body gazetteer, version 16, revision, 2015-02-19"
Vertical domain of dataset	identificationInfo[1]/*/extent/*/verticalelement/*/minimumValue identificationInfo[1]/*/extent/*/verticalelement/*/minimumValue identificationInfo[1]/*/extent/*/verticalelement/*/unitOfMeasure	Vertical extent of dataset • Minimum value • Maximum value • Unit of Measure	Vertical extent contained in the dataset. It will be provided as highest and lowest values with units of measurement Not mentioned on data submitter form, only on data repository form and later on public pages	Vertical extent contained in the dataset, provided as min, max value and unit of measurement. Complete when applicable. Element not shown on Submission Form Shown on data centre and publication form	DC	O	1	Real Real ISO/TS 19103 Subset P06 vocab
Spatial resolution	identificationInfo[1]/*/spatialResolution/*/distance (distance)	Spatial resolution • Spatial resolution value • Spatial resolution unit	Spatial resolution refers to the level of detail of the data set. 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	Complete when applicable Element not shown on Submission Form Shown on data centre and publication form	DC	O	1	Distance values UoM - SI units Subset P06 vocab

Temporal extent	identificationInfo[1]/*/extent/*/temporalElement/*/extent gmd:MD_Metadata/gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:endPosition/@indeterminatePosition == “unknown”	Temporal extent • Start date • End date	Time period covered by the content of the dataset. It will be provided as start and end dates.	Provide a start and end date to indicate the time period covered by the dataset	DS	M	1	The default reference system shall be the Gregorian calendar, withdates expressed in accordance with ISO 8601 (yyyy-mm-dd whereyyyy is the year, mm is the month and dd is the day)
Resource Temporal resolution	spatialRepresentationInfo/MD_Georectified/*/MD_Dimension/resolution/	Frequency of observations	Temporal resolution (mean period between each time steps) Not mentioned on data submitter form, only on data repository form and later on public pages	Indicate the frequency of time series, choosing from the list	DC	O	1	L03 – SeaDataNet Measurement Periodicity Categories
Date of publication	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date	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. 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	Not mentioned on data submitter form, only on data centre form and later on public pages Date that the original dataset and the completed data submission form are published and available to the public through the EMODnet portal	DC	M	1	The default reference system shall be the Gregorian calendar, withdates expressed in accordance with ISO 8601 (yyyy-mm-dd whereyyyy is the year, mm is the month and dd is the day)
Date of last revision	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='revision']/*/date	Date of dataset revision	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.	Element not shown on Submission Form Shown on data centre and publication form	DC	C Conditional, from perspective of processing or not in data centre Do not expose on xml	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)
Date of creation	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='creation']/*/date	Date of dataset creation	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	Indicate the date the package was created	DS	M	1	The default reference system shall be the Gregorian calendar, withdates expressed in accordance with ISO 8601 (yyyy-mm-dd whereyyyy is the year, mm is the month and dd is the day)

Specification	dataQualityInfo/*/report/*/result/*/specification dataQualityInfo/*/report/*/result/*/pass	INSPIRE Conformance • Specification • Degree	INSPIRE mandated element. Can be automatically generated as the submitter is supposed to have knowledge of INSPIRE specifications Three couples specification – degree will be completed in the backend: • 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 | Conformant • Corrigendum to INSPIRE Metadata Regulation published in the Official Journal of the European Union, L 328, page 83 | Conformant • Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services | Not conformant	Element not shown on Submission Form Shown on data centre and publication form. If Data Centre is aware of	DC Default: • 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 | Conformant • Corrigendum to INSPIRE Metadata Regulation published in the Official Journal of the European Union, L 328, page 83 | Conformant • Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services | Not conformant	M	1..*	Specification: • 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 INSPIRE Metadata Regulation published in the Official Journal of the European Union, L 328, page 83 Commission Regulation (EU) No 1089/2010 of 23 November 2010 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards interoperability of spatial data sets and services Degree • Conformant • Not conformant
Upload & Finish								
Uploaded file name	Ingestion Service specific	Filename	Filename of the data package submitted	Indicate the filename of the data package submitted with this form	DS	M	1	Free text
Process info								
UUID	Ingestion Service specific	Submission identifier (UUID)	Submission identifier automatically assigned by the Ingestion Service	-	DS AUTO			RFC 4122, ISO/IEC 9834-8:2005
Submission Status	Ingestion Service specific	Status	Status indicating the process stage of a submission	-	DS, DC, M	M	1	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

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 RFC1738 and IETF RFC 2056)
Date of publication	identificationInfo[1]/*/citation/*/date[./*/dateType/*/text()='publication']/*/date	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. 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	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, withdates expressed in accordance with ISO 8601 (yyyy-mm-dd whereyyyy 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 RFC1738 and IETF RFC 2056)
DOI	identificationInfo[1]/*/citation/*/identifier	DOI	DOI in case of academic publication	Identifier of the publication	M	O	1	