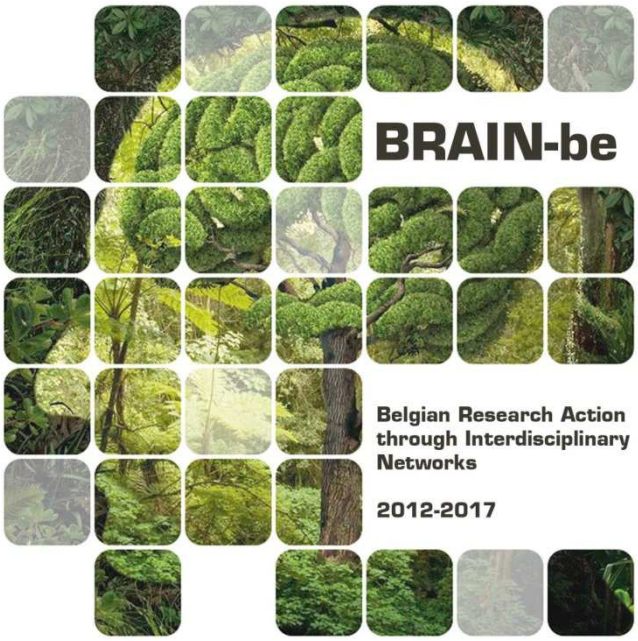


PAMIR Report on Metadata

T. Vandenberghe[1], Y. Christophe[2], R. Van Malderen[3], M. Lukach[3], S. Scory[1] and A. De Rudder[2]

[1] Royal Belgian Institute of Natural Sciences (RBINS)
[2] Royal Belgian Institute for Space Aeronomy (IASB-BIRA)
[3] Royal Meteorological Institute (RMI)

March 2018

PAMIR

A Portal to Atmospheric and Marine Information Resources

Brain-be Project BR/132/A6/PAMIR

Contents

Executive summary

The PAMIR project objective is to explore and develop generic concepts related to the description, documentation, classification and quality assurance of atmospheric and oceanographic data of the federal portfolio, in compliance with international standards and practices, including the INSPIRE Metadata Regulation. The purpose is to provide Belgian scientific institutes with an appropriate common framework to validate, document and publish their data holdings.

The work done in relation to metadata included the following three steps, respectively performed as tasks of Work Packages 3, 4 and 5.

- Review of the needs of the three communities involved, identifying common metadata fields and examining their compliance with INSPIRE.
- Tools integrating a metadata editor were envisaged, leading to the choice of GeoNetwork.
- The GeoNetwork catalogue and discovery service was integrated to the PAMIR project website and its metadata catalogue populated with a few test examples.

1. Introduction

The PAMIR project objective is to explore and develop generic concepts related to the description, documentation, classification and quality assurance of atmospheric and oceanographic data of the federal portfolio, in compliance with international standards and practices, including the INSPIRE Metadata Regulation [1][1]. The purpose is to provide Belgian scientific institutes with an appropriate common framework to validate, document and publish their data holdings.

The project Work Package 3 (WP3) deals with metadata, while WP4 and WP5 are respectively devoted to the implementation of a data catalogue and the provision of an oceanographic and atmospheric information portal integrating the catalogue as a feature. Work Packages 3, 4 and 5 are intertwined and were actually conducted simultaneously.

Three key milestones of the project are related to metadata:

- Design of a common metadata model suited to the documented scientific datasets and compatible with the INSPIRE Metadata Regulation [1].
- Design and development of a metadata editor based on the previous and compatible with the INSPIRE Interoperability Regulation [2].
- Set up of a centralized portal prototype offering a dataset catalogue (underpinned by the above metadata model) as well as online information on data validation and model evaluation.

Corresponding to these milestones,

- a first phase has been devoted to review the needs of the three communities involved, identify common metadata fields and examine their possibility to comply with the INSPIRE rules;
- after discussing the pros and cons of building up a brand new metadata editor versus adopting and updating an already existing tool, the project team picked the latter; tools integrating a metadata editor were listed and discussed, leading to the choice of GeoNetwork[2], which makes use of the INSPIRE ISO 19115 standard profile (Metadata for geographic information and services [W2]) and incorporates a discovery service;
- the GeoNetwork tool was integrated to the PAMIR project website [W3] and its metadata catalogue populated with a few test examples.

The communities involved in the project are the following.

Oceanographic science, represented by RBINS/BMDC [W4]
The oceanographic community has been using for almost a decade the OGC [W5] and the ISO 19100 standard series [W6] data and metadata. For example, two key European initiatives in the field, the Seadatanet project [W7] and the Copernicus Marine Environment Monitoring Service [W8] provide catalogues, metadata models and tools based on ISO19100 and refer explicitly to INSPIRE compliance as a main target for their metadata and services.

[1] Numbers in brackets refer to Section 7 (References). Numbers in brackets preceded by the letter *W* refer to Section 8 (Web sites and downloadable documents).
[2] GeoNetwork is an open source catalogue application to manage spatially referenced resources [W1].

Atmospheric and climate science, represented by IASB-BIRA [W9]
Atmospheric and meteorological science, represented by RMI [W10]
The situation for atmospheric, meteorological and climate sciences is varied. Different standards and common practices coexist for the data format, supported metadata format and catalogue software, with conversion tools between many of the formats. The prominent format and metadata conventions in the climate community are the NetCDF format [W11] and the supporting Climate and Forecast (CF) Conventions and Metadata [W12]. NetCDF is essentially, though not exclusively, designed to record gridded data. The CF Conventions mainly address geophysical variables depending on longitude, latitude, height (altitude, air pressure, potential temperature) and/or time. CF provides standard names for a continuously updated list of variables.

2. Metadata: generalities

2.1. Definition

The term metadata (data about data) appeared in the nineties with the plan to describe and manage internet resources, but metadata existed and had been in use for a long time before the word was invented. Wherever numerous objects of a same kind need to be discriminated, metadata are handy. The term may be essentially understood as a synonym of “description”. Any field of study requiring classification and cataloguing of objects or ideas secretes its own system of metadata. Typical examples of metadata generators and consumers are archives: every library relies on some metadata scheme that makes the search for a document possible within a catalogue raisonné where books are listed and briefly described, usually in some coded fashion. Consulting a catalogue implies the prior knowledge of its underlying convention – that is, of the metadata scheme. Hence the interest of defining metadata schemes that are common to the collectivity of actors and users of a given domain, which then become standards. A worldwide example is the International Standard Bibliographic Description (ISBD) for written documents [3], produced by the International Federation of Library Associations and Institutions (IFLA) [W13].

The advent of digital information has brought an additional dimension to metadata to the extent that they can be – and very often are – automatically produced and read, and are usually made searchable via online interactive catalogues. Institutions and professional activities that generate data and for which sharing information is crucial are confronted to a similar issue. In Earth Observation sciences, the production of huge amounts of satellite and model data has made the standardisation of metadata indispensable.

2.2. Scientific metadata

It is a truism to state that science rests upon observation. A large part of scientific activities consists in the analysis of data that translate observation or calculation results in some coded way, often (but not always) numerical – and, today, generally in electronic format. The data sets most commonly encountered in Earth physics are records of numerical values of some physical quantity, usually derived from direct measurements of some other physical quantity through an algorithm and, in the best case, provided with an evaluation of the associated uncertainty; the data may also be the outcome of some numerical encryption of non-numerical observations; they may stem from evidence-based speculative attempts to reproduce or forecast the evolution of some aspect of the physical world using computer models.

A scientific dataset is composed of the values assumed by a set of variables (the “dependent” variables) over some domain of values assumed by some “independent” variables. The data themselves can range from raw measurements, processed data, continuous data, consolidated data, modelled or simulated data and can appear in continuous or rasterized (in 2 or 3 dimensions) form. Data in an oceanographic or atmospheric context have a temporal and/or geographical extent and can be sliced by an arbitrary temporal or geographical extent; these data subsets (data series when subset by time) need metadata as well and may be generated in a dynamic way (so

that metadata should be dynamically generated as well). (Dis)aggregation drives granularity and can occur on other levels as well, ie. mission, research platform, sample, individual observation,… Furthermore the data content itself may be dynamic, e.g. a dataset can be appended continuously or data points can be modified for whatever reason. Metadata is critical for automated data citation, data auditing and data versioning purposes.

In any case, in order to infer meaningful information from the data, the analyst must not only know how to "decode" the data, but also the data format and ancillary information of various nature, such as the data unit, observational conditions or computing hypotheses and approximations, references of papers describing the retrieval algorithm, version number and changes with regard to the previous version, who to contact, and, last but not least, a documented assessment of the uncertainty (or uncertainties) affecting the data. Even when the observer (or the data collector, retriever or modeller) and the analyst are the same person, which is less and less commonly true, this knowledge must be stated explicitly to enable other persons to perform possible future re-analysis. Metadata ensure that this knowledge is shared and passed on seamlessly to anyone willing to use or re-use the data.

2.3. Data and more

Metadata imply the existence of data. Data are not necessarily numbers, but can be any elements carrying information: books, internet resources, etc. By extension, the term metadata is being more and more often applied to objects of various natures.

In INSPIRE, metadata may document not only *datasets* and *dataset series*, but also *services*, all called by the generic name *resources*. In the INSPIRE realm, a *service* belongs to one of the following four categories of *Spatial Data Services* (SDS): Discovery, View, Download and Transformation [4] – the notion of "Invoke" SDS (service that, given some data input and a chain of other services, produces a data output), initially also considered by INSPIRE, now seems to have been abandoned.

This evolution appears from the fact that many metadata elements of a dataset are applicable to the service delivering the data as well, and furthermore, that these corresponding elements could even share the same value and multiplicity (in which case they are redundant). Services are primarily data delivering services (discovery-view-download), but there is a trend from data delivery towards aggregation, visualization and even embedded model workflow tools (examples: CoESRA virtual experiment environment[3], eReefs[4]). These (sub)services may either be embedded within the same service or be external. Exactly how these services are chained can be described in metadata (of the data product at a chain stage or of the data service) as well.

The PAMIR project does not address service metadata, but rather focuses on metadata describing datasets *stricto sensu*.

[3] https://espace.library.uq.edu.au/view/UQ:345046
[4] https://research.csiro.au/ereefs/

2.4. Granularity

The fact that metadata refer to data sets prompts the question of what demarcates a data set (or dataset). The answer will be a matter of context and the nature of the criteria used to delimit a dataset will vary. For example, in Earth observation (EO) sciences, a dataset may be defined as the set of data collected at a common site or from a common platform, included in a common file or in some defined series of files, etc.

One metadata (sub)set may refer to a datum, to a dataset, to a series of datasets… In the CF Conventions, for example, the metadata are divided into "global attributes" pertaining to the dataset as a whole (e.g. *title, institution, history,* …) and information about coordinates, variables and their specific attributes like unit, valid range, filling value (to signal gaps in the data), standard error, reference frame, etc.

2.5. Functional types

Metadata are essential to enable potential users to locate, identify and access the data they need, to assess their fitness for purpose, understand their meaning and limitations and actually use the data in particular applications. Their function (i.e. supporting data use) explains the distinction often made between various types of metadata, depending on which stage of data usage they relate to. Many metadata schemes include three basic subsets:

i. elements informing on the existence, location and availability of the data, making the search for – and, but not mandatorily, the access to – the data possible ("discovery" or "high-level" metadata), usually through some catalogue – e.g. data set name, type of resource, brief description, main parameters included in the data, dataset index in some list or classification system, author's or provider's name or other responsible party (usually an institution or organisation rather than a person), contact, associated scientific project, date of issue, URL addresses, conditions of access and use, …;

ii. elements describing the data formal display (or arrangement of the data in the dataset), providing guidance on how to read the data once acquired or accessed (whether the reading is done in some automated way or not) – e.g. language, data electronic format, number of columns in a data file, which coordinate or datum is recorded in which column, how data are aggregated into chunks, …;

iii. elements pertaining to the data nature and processing, allowing their understanding and informed use – e.g. provenance; conditions of data collection; physical processes involved; successive stages of the data processing chain; algorithm characteristics; nature, unit and uncertainty (or other quality indicators) of a physical quantity; codes used for to flag particular data; "fill" value used to signal missing or suspicious data; warnings and limitations to take into account in applications; paper references; …

Any of these types may be split up into finer categories if so wished. It should be noted that the three types identified above actually overlap, so that frontiers between them are somewhat fuzzy.

There are circumstances where some data may be inaccessible – e.g. because they are recognised confidential by law or by individuals – or accessible only to authorised users or after payment of some fee, and still be described by public metadata. The data targeted by the INSPIRE

Directive [5] must be made publicly and freely available (possibly with applicable conditions of use) via some dissemination service. The PAMIR project is not concerned with the distribution of the data, but only with their documentation via some metadata catalogue (including pointing to the data distributor).

2.6. Metadata schemes

As mentioned above, metadata are elements of description of a defined category of objects – considered as data in a very general sense. These descriptive elements or "fields" are usually structured (they are divided in categories or chapters, they relate to each other, they obey certain rules). The generic elements and their organisation form a conceptual metadata model, or scheme, or system, or profile – in this context, the terms are synonyms. Since the scheme reflects common properties of the objects it encompasses, the broader the variety of the data, the more laconic the metadata scheme

Standard profiles of metadata have been developed for certain narrowly or broadly defined types of data. One of them is associated to the European Directive INSPIRE [5] applying to any dataset in electronic format, relating to geo-located objects and detained in a Member State of the European Union (EU). The categories of data sets or rather the fields of interest targeted by the INSPIRE Directive are listed in its three annexes. The INSPIRE Metadata Regulation [1] defines the minimal discovery metadata required for a dataset to be integrated to the INSPIRE data network. This scheme is summarised in Appendix A.

2.7. Metadata quality

Metadata only contributes to the service and product quality on the condition that the metadata themselves meet certain quality criteria.

"Good" metadata are

- specific (addressing the level of detail required to an in-depth understanding);
- accurate (not leaving dark or fuzzy areas, ambiguities or room for misunderstanding);
- intelligible (formulated in a logically and grammatically correct language);
- explicit (avoiding coded information, abbreviations and undocumented conventions, and providing references where needed);
- complete (exhaustive enough to allow data discovery, assessment of fitness for purpose and proper use);
- consistent (organised and classified according to some sensible scheme, both of which ease the metadata search).

3. Metadata for georeferenced data

3.1. Starting point

Atmospheric and climate data

As stated in the introduction, data formats and metadata schemes used in atmospheric, meteorological and climate science are numerous. Rather than picking particular examples, some generic theoretical scheme was compared to the INSPIRE scheme described in Appendix A in order to assess the capacity of the latter to capture key features of atmospheric data. This theoretical scheme was elaborated in the context of the EU FP7 project PASODOBLE [6] and is therefore called “the PASODOBLE metadata scheme” hereafter.

The conclusion was that, although always possible to derive INSPIRE-complying metadata from the PASODOBLE metadata, key atmospheric information would be lost or modified in the process. The most important points regard data quality and dimensions.

Data quality. In science, the assessment and provision of the data uncertainty (or other quality indicators informing on the probable “closeness to truth” of the data), even if not its only constitutive element, is at the heart of the concept of quality, as defined for example by the Quality Assurance Framework for Earth Observation (QA4EO) [W14]. INSPIRE, on the other hand, understands quality mainly as a composite of the geographic resolution of the data and the degree of formal compliance of the metadata with the INSPIRE scheme. Beside these two distinct metadata fields, a third field called “lineage” allows to provide some general statement on data processing and data quality (for example, whether the data have been validated), but this information remains very vague and stands on the same level as a number of unrelated pieces of information, such as whether the described dataset is the “official” version of it.

Dimensions. For atmospheric or oceanographic data, the vertical and time dimensions are crucial. With its geographic focus, INSPIRE dwells on the two horizontal dimensions but never explicitly considers height, depth and time. It is not impossible, in the INSPIRE discovery metadata, to mention the dependence of the described variable on the vertical and on time, but it is embedded in some more general metadata field such as the resource abstract, where the user will not necessarily have the curiosity to look for such a piece of information.

In conclusion, when tackling the question of the metadata editor, a tailor-made tool should be built to do justice to atmospheric data, or some existing tool could be adopted at the condition of being flexible enough to allow the creation of new metadata fields appearing in the associated catalogue or search engine with the same importance as the INSPIRE fields do.

Oceanographic data

As already mentioned in the introduction, the oceanographic community has been using for almost a decade the OGC [W5] and the ISO 19100 standard series [W6] data and metadata. For example, two key European initiatives in the field, the Seadatanet project [W7] and the Copernicus Marine Environment Monitoring Service [W8] provide catalogues, metadata models and tools based on

ISO19100 and refer explicitly to INSPIRE compliance as a main target for their metadata and services.

3.2. Metadata standards and profiles for georeferenced data

Metadata standards provide metadata profiles that can be consistently and meaningfully applied over multiple datasets. They may also provide an implementation of this profile in the form of a specific scheme. They establish a proper understanding of the meaning (semantics) of datasets and the disparity of the meaning between datasets. In general, metadata frameworks provide guidance on general or specific aspects of metadata.

INSPIRE

The European Directive 2007/2/E, establishing an Infrastructure for Spatial Information in the European Community (INSPIRE) [5] was published in the Official Journal of the European Union (EU) on the 25th of April 2007. The Directive scope includes any dataset with a social usefulness, which possesses a geographic extent or is related to some geographic location (which is the way "spatial" must be understood in this context). Its double objective is to harmonise such datasets over the continent and to make them available to the community (citizens, administrations, agencies, universities, research institutions, hospitals, doctors, policy makers, etc.) all over Europe, through a network of information services.

The "data themes" targeted by INSPIRE are listed and grouped in three annexes to the Directive. Items 7, 13, 14, 15 and 16 of Annex III are respectively entitled "environmental monitoring facilities", "atmospheric conditions", "meteorological geographical features", "oceanographic geographical features" and "sea regions". Additional INSPIRE data themes are tightly bound to research in ocean and atmosphere sciences, either because of their feedback role in the Earth physical system or as being directly influenced by marine or atmospheric phenomena, including climate change : land cover (II.2), land use (III.4), human health and safety (III.5), agricultural an aquaculture facilities (III.9), population distribution – demography (III.10), natural risk zones (III.12), bio-geographical regions (III.17), habitats and biotopes (III.18), species distribution (III.19), energy resources (III.20).

The INSPIRE Directive itself does not include any practical consideration regarding data sets, metadata or information services. These are dealt with in INSPIRE subsequent Implementing Rules (IR) which have been established by the INSPIRE drafting team after undergoing some review by the INSPIRE registered Spatial Data Interest Communities (SDIC) and Legally Mandated Organisations (LMO). European Regulations or Decisions have been published about top-level (or "discovery") metadata, monitoring of INSPIRE implementation and reporting by the member states to the European Commission (EC), network services, data and service sharing and various aspects of interoperability.

Following Article 5 of the Directive, the EC has also published the Commision Regulation No 1205/2008 with regards to metadata [1]. As described above, also an IR on metadata was published, the "Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (MIR)" [7]. This document lays down rules for spatial data sets, spatial data set series and spatial data services. The set of metadata elements laid down in 1205/2008 [1] is quite a minimal set (see Appendix A), and one could argue insufficient or even misleading for more

advanced fitness-of-purpose enquiries. For example, in the metadata regulation, spatial resolution is presented as a quality indicator without much other context (see previous section). When mapping to ISO 19115 however, spatial resolution is a part of identificationInfo, a 'neutral' element that (like any other) can still contribute to an assessment of fitness-for-purpose.

The former two data specifications focus more on how to expose the data via services in the OGC O&M scheme (i.e. ISO 19156) than on the metadata itself, while the latter focuses on describing geographical features themselves. The "Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE Annex II and III data specification development" (2014) form a valuable resource in this respect.

Finally, technical data implementation specifications have been published by thematic working groups (TWG) for all INSPIRE data themes. These guidelines, making use of international standards, are not legally binding. There is one specification per INSPIRE data theme, with the exception of the atmospheric and meteorological data themes, which have been merged. Data specifications relevant for this project encompass:

- "Atmospheric conditions and meteorological geographical features"
- "Environmental Monitoring Facilities"
- "Sea regions"
- "Oceanographic geographical features"

ISO 19115 and related geographical data standards from ISO

The original ISO 19115:2003 standard on geographic information - metadata dates from 2003. An XML schema implementation was published in three years later in 2006 in the ISO 19139 standard. ISO 19115-2 was created in 2009 to accommodate imagery and gridded data. The most recent version, 19115-1, was published in 2014 and an XML implementation soon followed in May 2015. This implementation, 19115-3, is the successor of ISO 19139.

The ISO 19115 standard states that individual communities may develop a "profile" of the International Standard according to best practices and consensus within the community of a certain field of application. A profile can be implemented as a template that restricts or constraints the fields/elements a user can see with a set of validation rules to check compliance; codelists, codelist entries (vocabularies), or changes in optionality and multiplicity can be created. On top of this, new metadata elements or entities can be created. A profile can also include an agreement on the level of data granularity, ie. what defines a dataset, a dataset collection, as this can simplify the metadata model.

Known ISO 19115 profiles are the SeaDataNet CDI profile, the World Meteorological Organisation (WMO) Core Profile, the INSPIRE profile, etc.

SeaDataNet CDI Metadata profile

The Common Data Index (CDI) metadata format is a marine profile of the ISO 19115-19139 spatial metadata content standard developed by SeaDataNet. SeaDataNet is a standardized system for managing the large and diverse data sets collected by the oceanographic fleets and the automatic observation systems, and operates across European institutes. The latest version, 10, dates from

September 2013. It provides an index to raw data on individual samples, cores and measurements and a unique interface to access these data sets online, the CDI Data Discovery and Access service. It covers physical, geological, chemical, biological and geophysical data, acquired in both European and global seas. The CDI format can be considered a de-facto standard for marine metadata in Europe. More than 1.7 million CDI entries are currently available. Datasets in the CDI service have a small granularity, mostly collecting multiple variables at sample level. Examples that correspond to a single CDI entry: a single sediment grab, a single CTD cast; a (long) timeseries of a single instrument at a single station; a single track made by a single instrument in a seismic/bathymetric survey, each possibly covering multiple parameters.

SeaDataNet has recently set-up a web service with virtual collections of CDI metadata entries. The CDI entries are aggregated by discipline, data centre (based on controlled vocabularies), and geometric type (point/track/surface) which result in about 480 CDI collections. These virtual collections have no metadata attached except the metadata of the individual granules.

The XML encoding has been upgraded to the 19139 Schema, which is adopted as part of the EU INSPIRE Directive Implementing Rules. During this extension, the CDI schema has also been made INSPIRE-compliant.

Broadly, the index provides:

- The type of parameters it refers to
- The location and time of the measurement
- The method to retrieve the information
- Where applicable, basic preview (track)

During its history, it has been upgraded by Geo-Seas. Geo-Seas is a Pan-European infrastructure for the management of marine geological and geophysical data. Some of the improvements made available were:

- GML objects to support tracks and polygons next to points
- Service bindings for extra services (e.g. previewing)
- Resolution (spatial / temporal)
- Multiple instruments (to be able to include the positioning systems adjacent to the primary measuring device)

CDIs are created on a service/project basis: if a project of a service (i.e. research institute, data centre/NODC) has new data for which no CDI index exists yet, CDI indexes are created for each new sample in an automated fashion. For the discovery service, this has been done in a retroactive way. To access the data, either NetCDF files are available, or localized SQL views are created that contain all the data plus the CDI index. This index is a concatenation of service+project+sample. For each unique index in the view, dedicated software generates one CDI xml file that describes the metadata for this sample (because of this, the bounding box of one CDI metadata file is actually a point). Together with the CDI file, a mapping file points the column names coming from a controlled vocabulary to the local column name used in the view, and a coupling file registers the NetCDF file or the corresponding query from the SQL view.

Within the context of the PAMIR project, the CDI schema has been chosen to compare it to PASODOBLE as it is the most extensive schema in the SeaDataNet metadata schemas (ie. CSR, CDI, EDMED, EDMERP and EDIOS). However, the granularity of the datasets the CDI format points are rather small and don't really correspond to a dataset in the data product sense.

SeaDataNet EDMED Metadata profile

The European Directory of Marine Environmental Data (EDMED) is a comprehensive reference to the marine data sets held within European research laboratories, data centres/NODCs. These datasets are the continuous data output of individual projects from their start to end. Just as CDI, EDMED has a metadata profile based on ISO 19115. CDI datasets are granules, the smallest data elements that make sense in an ocean science context. These CDI datasets can be aggregated to data series, which share or summate the common characteristics of the encompassed datasets and are themselves datasets. These data series might or might not be datasets in the EDMED context.

Remarks on thesauri

- INSPIRE mandates to properly cite the thesaurus or ontology when used; for some metadata, INSPIRE constraints to codelists defined in ISO 19100 standards and recommends usage of GEMET thesaurus (http://www.eionet.europa.eu/gemet/)
- Usage of controlled vocabularies greatly varies between (sub)communities (from none to rigid or non extensible, ...); the definition of the controlled vocabularies may or not follow RDF or SKOS standards, follow a versioning scheme, etc.
- See also e.g. the ESA project 'PROD-TREES': http://wiki.services.eoportal.org/tiki-index.php?page=Prod-Trees+Project

3.3. PAMIR approach

As a result of the discussion that took place within the PAMIR working group, an ideal set of common metadata elements has been identified for marine, atmospheric and meteorological datasets. These metadata elements are not all present in the different metadata standards/profiles available.

As regards metadata, during the course of the PAMIR project, the following activities have been performed:

- Identify necessary metadata elements based on reference metadata models of each community
- Perform a mapping between models
- Explore general suitability of ISO 19115
- Describe shortcomings of ISO 19115 for selected datasets

Necessary metadata elements

In order to be useful for multiple scientific communities, a shared metadata model like PAMIR should have a set of core characteristics that fit to the practices and standards of each community.

The metadata model developed in the framework of PASODOBLE is a metadata model proposal for products distributed by local air quality services initially developed in the context of GMES (predecessor of Copernicus). Although improvable, it provides many of the wished-for characteristics. It is particularly suited for particular Air Quality products and has been designed with INSPIRE compliance in mind. Because of its relative ease to map to ISO 19115 (and hence to the CDI profile), the PAMIR collaborating institutes consider the PASODOBLE metadata model as a valuable model to record oceanographic metadata as well.

From the PASODOBLE metadata model, the following essential generic characteristics of a metadata model have been identified. For the atmosphere community, a metadata model should:

- Operate on data products (or data resources), data services and optionally on data models.
- Be able to separate and link together data series and datasets.
- Describe the domain of definition in 4-dimensional extent.
- Describe spatial and temporal representation information:
    - representativeness: point/gridded; instantaneous or averaged over time
    - resolution
    - boundaries
- At least be product-centric. Service metadata should either be separately described or embedded in a data product, not (only) the other way around
- Have a detailed set of data quality indicator elements, flexible enough to express adherence to preset validation protocols (e.g. QA4EO) on a dataset level (this high level seems appropriate for discovery but may not be for usage purposes). Be able to express quality indicators separately for each dependent variable. For individual quality indicators, be able to indicate confidence interval and level.
- Have a detailed set of lineage elements to describe every process step of the data product, including quality indicator elements.
- Have conformity to one or more references (for data and services).
- Have use and access constraints.
- Make use of international metadata standards and frameworks, notably INSPIRE.

Additionally, in the marine context (taken from CDI) the following characteristics are needed:

- Support trajectories (tracks, lines)
- Constrain freetext elements to the SDN thesaurus:
    - keywords:
        - devices, platforms, parameter discovery codes, programs (EDMERP), ports, countries, platforms, waterbodies and Marsden codes
    - code for identifiers: EDMED code, CSR code, CRS code (MD_Identifier/code)

- formats for distribution (MD_Format/name)
- institute code for responsible parties: EDMO code (CI_ResponsibleParty/organisationName)
- country code for addresses (CI_Address/country)

Data granularity

In order to speak of a data delivery service, it is necessary to mention the granularity. The question of the most practical level of data granularity is a function of:

- The technical difficulty of dividing a dataset into its granules and applying the same scheme to other datasets. When datasets differ too much in their lineage (other origins, types and derivation processes), and without automation and infrastructure, this is impossible.
- The flexibility in combining that a small granularity provides. The to be returned data granules are constrained beforehand on the basis of one or more parameters and can be aggregated later on the basis of other parameters, unknown to the original data provider.
- The technical difficulty and overhead of (either manually or automatically) joining data granule results.

Once operational, the PAMIR portal will contain datasets of very disparate granularity, type, origin and lineage. PAMIR is intended as a dataset repository service, not as a dataset recombination service in the spirit of CDI, where strict convergence is needed. In the SeaDataNet Common Data Index workflow, datasets are reduced to the sample level granule, which provides a lot of flexibility.

3.4. Compatibility and mapping

Concordance between PASODOBLE and the INSPIRE Metadata implementing Rules

The PASODOBLE model has been developed with compliance to the INSPIRE metadata regulation in mind – although the inclusion of specific requirements (see Section 3.1) makes the mapping between the two model (proposals) challenging. This challenge seems resolvable if one considers that the INSPIRE Metadata implementing Rules impose the use of the ISO 19115:2003 metadata scheme, which is a metadata standard having a remarkable completeness. INSPIRE compliancy is still attainable in complex metadata records that make use of the full suite of possibilities within ISO 19115.

The shortcomings (reported in Section 3.1) of the metadata elements in the table of the Annex Part B are alleviated to a satisfactory extent. The main critique is that the PASODOBLE metadata model asks for quality and dimensional information about each separate variable (independent and dependent). We have compared the PASODOBLE metadata model with the ISO 19115 norm and have found solutions for this critique. What seems to be difficult to achieve in the ISO norm is information on dependent variables and more specifically quality indicators separately for each dependent variable.

With regards to Data quality, the ISO 19115:2003 norm identifies two quality descriptions, i.e. DataQuality elements and Lineage elements. A metadata file can have zero or more DataQuality elements that can describe quality and history of a dataset in a quantitative fashion or in a

descriptive form (with the Lineage element). Lineage describes the source data and the processes applied; multiple process descriptions are separate instances. DataQuality offers a rich set of elements to quantify completeness, consistency, scope, topology, positional accuracy etc. Both lineage and data quality are characteristics of the whole dataset and can't be applied to the individual variables (independent or dependent) of the dataset, although they can be textually mentioned.

PASODOBLE demands clear information on Dimensions of independent and dependent variables. This can be done in the ISO 19115:2003 norm for the independent variables location, depth/height, time and path in the axis by using the axisDimensionsProperties field. ISO 19115 has classes for describing the independent data parameters of coverages (geospatial information representing space/time-varying phenomena, usually gridded) and features (discrete objects represented as points, lines and polygons) in the ContentInformation package. This package is complicated to use and because of this the Australian Ocean Data Centre has developed the Marine Community Profile (MCP)[5]. This includes a dataParameters element for the Data identification instance.

Concordance between ISO 19115/CDI and PASODOBLE

Identifying the concordance between the metadata models commonly used or proposed as a standard is an essential first step in developing a common scheme or at least a common usage of an existing scheme. We have compared all the elements of the PASODOBLE metadata scheme with the ISO 19115/CDI scheme (one-way crosswalk). We have worked out a table-based crosswalk and a textual mapping. In the table-nased crosswalk, elements can be mapped absolutely, relatively or not at all (non-common fields). The fields that are not common between both models have been discussed in a textual mapping and in some cases a solution to a mapping has been found. When no mapping with the CDI profile was possible, we have searched for a mapping with the generalized ISO 19115:2003 standard. In some cases, we explored the possibilities of the ISO 19115:2014 version. The general result is that still a significant portion of the PASODOBLE fields (<20%) can't be mapped in the ISO 19115 standard. On the other hand, the ISO 19115 standard has other fields not present in the PASODOBLE metadata model, and it has a well though-out hierarchy. The necessity of these PASODOBLE fields should be further examined.

The textual and table-based mappings can be found in Appendix B.

[5] The AODC's Marine Community Profile uses the data parameters class to model parameter information: Profile http://mcp-profile-docs.readthedocs.io/en/stable/extensions/dataParameters/index.html

4. Catalogues

Scientific data repositories may be provided with (web) catalogue(s) that list the available datasets and provide metadata for them. Web catalogues and repositories may exist in one service location or in isolation of each other (in which case they can be linked by broker systems). Repositories of repositories can serve as metacatalogues.

Catalogues can be implemented using different paradigms; for example, dedicated software uses defined protocols to serve metadata and redirect to data repositories (e.g. GeoNetwork implements the CSW protocol for geospatial data, THREDDS defines a catalogue protocol for geoscience data repositories). The use of defined protocols allows to harvest metadata and hence (with appropriate toolset and configuration), to build automated custom (meta)catalogues.

Catalogues have an accompanying ruleset:

- what underlying metadata model and profile is used; often a generic framework is provided that can operate on any profile of the model.
- what enters the catalogue, i.e. which level of granularity defines a dataset, a dataset collection
- how metadata which are already available from the data repositories themselves can be ingested in the catalogue (depending on the granularity and the system architecture). This relates to what extent the data is separated from metadata. For instance the SeaDataNet Common Data Index catalogue makes use of a distributed broker architecture that retrieves data requests (driven by metadata stored in the catalogue) from the NODCs.

Catalogues are generally backed by some metadata editor functionality (see next section) in which case it is better to speak of metadata tools. These generally have an interface separating front visitors from back-end users.

4.1. Examples of GeoNetwork-based catalogues

Finnish Meteorological Institute

- URL: http://catalog.fmi.fi/geonetwork/srv/en/main.home
- based on GeoNetwork 2.6.4 ?
- mainly in Finnish, with some translations to English
- ~4550 records
- data (relevant to PAMIR): atmospheric (many datasets, including radar)
- metadata in ISO 19139/ ISO 19115, with data quality metadata claiming conformance to INSPIRE, e.g. D2.8.III.13-14 Data Specification on Atmospheric Conditions and Meteorological Geographical Features - Technical Guidelines

Nationaal Register (The Netherlands)

- URL: http://www.nationaalgeoregister.nl/geonetwork/srv/eng/search

- based on gonetwork 2.10 (?)
- english and dutch
- ~8000 datasets and ~500 services (~8500 records)
- data (relevant to PAMIR):
  - oceanographic (e.g. NIOZ Royal Netherlands Institute for Sea Research)
  - atmospheric (many datasets, including radar, ozone measurement, etc. from KNMI)
  - metadata in ISO 19139/ ISO 19115, but apparently no specific INSPIRE data quality references

4.2. Examples of THREDDS-based catalogues

http://data.nodc.noaa.gov/thredds/catalog.html

http://gsics.eumetsat.int/thredds/catalog.html

http://motherlode.ucar.edu/thredds/catalog.html

4.3. Informal catalogues

Catalogue without an identified protocol: e.g., the MACC catalogue (http://www.gmes-atmosphere.eu/catalogue)

5. Metadata editor

Metadata editors allow users to edit metadata elements interactively without needing to resort to editing actual XML files. They also allow creation and use of new metadata templates and can provide consistent use of vocabularies over the different datasets. They are often embedded in a metadata management tool.

5.1. PAMIR approach

The partners' datasets are not all in NetCDF, some might be databased, as is especially the case for BMDC. As the CF Convention is not universally used over all datasets, a THREDDS implementation has been considered but is unpractical. THREDDS Data Server (TDS) is an powerful environment that can store and deliver both the data and metadata of NetCDF, OpenDAP, and HDF5, binary datasets, OGC WMS/WCS and can provide automated metadata analysis and ISO metadata generation.

GeoNetwork is one of the few metadata catalog systems that is capable of harvesting THREDDS catalogs, although technical difficulties have been found doing this in test runs.

During the project, the project partners realized that the ISO standard provides the most flexibility and from this viewpoint a metadata editor/portal built around the ISO standard is the most feasible solution.

After discussing the pros and cons of building up a home-made metadata editor versus adopting and updating an already existing tool, the PAMIR project team picked the latter, with the main arguments in favour of this choice being

- the fact that its default metadata profile is INSPIRE-compliant;
- its flexibility (possibility to expand the default metadata profile with self-defined metadata fields, thus getting round the obstacles signalled in Section 3.1);
- its generalised use in the oceanographic community;
- the fact the Belgian federal infrastructure for spatial information (www.geo.be) has mandated its use to harvest the instances of the federal scientific institutions.

It was thus decided to implement a GeoNetwork instance for PAMIR [8]. This was done at IASB-BIRA, first as a test, than as an online catalogue integrated to the PAMIR website [W3]. The project did not have the resources to work on the possible expansions of the GeoNetwork default metadata profile to meet the requirements listed in Section 3.1. Instead, each partner entered a few test datasets to see whether the tool delivered all basic needs.

As no INSPIRE checker is (currently) available on the instance, the metadata have been manually verified to be INSPIRE-compliant. The xml of some datasets have been validated in the INSPIRE geoportal (http://inspire-geoportal.ec.europa.eu/validator2/).

On the long term, a promising avenue is a distributed approach whereby PAMIR operates as a metacatalogue: a) The PAMIR portal harvests from the separate metadata catalogues or

repositories of the two institutes, which can be a mixture of GeoNetwork or THREDDS instances; b) the federal INSPIRE Geoportal currently in development by the NGI/IGN harvests from the institutes own node. In retrospect, we think a distributed approach would have provided the most added value for the PAMIR project partners.

5.2. GeoNetwork implementation tests

GeoNetwork is a metadata catalogue web application that allows metadata editing and searching based on ISO 19115, although it can be set up with multiple metadata standards such as DublinCore. It is possible to fork GeoNetwork and change the editor for a specific scheme or profile; this requires writing a set of xslt transformations. As mentioned above, it is currently used as is in PAMIR, since the project had no resource to set it up with new profiles. But it is recognised by the project partners as an ideal tool if adequate profiles can be added.

Use cases highlight deficiencies of (INSPIRE-compliant) ISO 19115 profile provided with GeoNetwork.

Practical concordance between ISO 19115 and a set of datasets, making use of GeoNetwork

Datasets entered:

Atmosphere science

- RMI Probability of hail from Wideumont radar (from data in the spreadsheet by Maryna Lukach)
- BISA MAXDOAS vmr vertical profile data of NO2 at Xianghe (China) (from GEOMS metadata provided by Caroline Fayt, part of NDACC measurements)
- BISA BASCOE NRT analyses of stratospheric species using MLS observations
- BISA BASCOE NRT analyses of stratospheric species using MLS observations (xml metadata according to WMO core profile collected from MACC catalogue)
- NDACC stratospheric chemical species and parameters (contains the BISA MAXDOAS dataset) (from NERC data catalogue, http://data-search.nerc.ac.uk/search/full/catalogue)

Ocean science

- Multibeam track MB1002_FOD_BE_20060302T1957_KBMB (Common Data Index-level)
- Belgica cruise 2006/04A (Cruise Summary Report-level)
- Real-time data of cruise BE2014/22 (Cruise Summary Report-level)
- Bathymetric surveys of Western Scheldt and river Scheldt (Belgium) since turn of the century (dataset level)

Bathymetric surveys

Dataset taken from the EDMED directory[6]. This dataset is collated by BMDC but is not part of its data holdings. In total BMDC has published 97 datasets in EDMED. Given the quantity of datasets it is quite difficult to gather many details beyond what's needed in EDMED about the dataset (asking each isn't really the point). This example illustrates that it will sometimes be impossible to

[6] http://www.bodc.ac.uk/data/information_and_inventories/edmed

provide INSPIRE-compliant metadata. The dataset doesn't parse ISO 19115 because no distributionFormat can be given.

- Collating centre taken as role: processor
- can't constrain organisation from thesaurus list (see higher)
- Used unique resource identifier for Local identifier and Global identifier; used CI_Citation to point it to processor and distributor respectively
- Maintenance frequency: unknown
- Keyword type: theme
- Access constraints: other restrictions; use constraints: IPR; other constraints: data access by negotiation
- Character set: deleted as unknown cannot be chosen
- geographic extent only described in description text and in Geographic element/Geographic description/Identifier "Western Scheldt and Zeeschelde, Belgium".
- Start date: 19th century: 1800-01-01; Indeterminate position (unknown)
- End date: Indeterminate position (now)
- Data quality with scope dataset: lineage statement taken from SDN CDI: 'The data holding centre applies standard data quality control procedures. Ask the data holding centre for details.'
- INSPIRE-compliant datasets require a positive conformity statement. It is advised to use a DQ_DataQuality.report.DQ_ConformanceResult with values
    - specification.CI_Citation.title.CharacterString = ‘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’
    - specification.CI_Citation.date.CI_Date.date.Date = ‘2008-12-04’
    - specification.CI_Citation.date.CI_Date.dateType.CI_DateTypeCode.@codeListValue = ‘publication’
    - explanation.CharacterString = ‘See the referenced specification’
    - pass=false (in this case)
- Distribution:
    - File format unknown: cannot be set
    - Ask file by emailing: set as description in online resource as this seems only place to describe this (actually there is no online resource)

Elements not set:

- character set

- vertical element
- spatial representation info
- distribution format

INSPIRE checklist:

1.1	Resource title	1	ok
1.2	Resource abstract	1	ok
1.3	Resource type	1	Not explicitly
1.4	Resource locator	0..*	**ok**
1.5	Unique resource identifier	1..*	ok
1.7	Resource language	0..*	ok
2.1	Topic category	1..*	ok
3	Keyword	1..*	ok
4.1	Geographic bounding box	1..*	Not ok
5	Temporal reference	1..*	ok
6.1	Lineage	1	ok
6.2	Spatial resolution	0..*	ok
7	Conformity	1..*	ok
8.1	Conditions for access and use	1..*	ok
8.2	Limitations on public access	1..*	ok
9	Responsible organisation	1..*	ok
10.1	Metadata point of contact	1..*	ok
10.2	Metadata date	1	ok
10.3	Metadata language	1	ok

7. References

1. EC, Commission Regulation (EC) N° 1205/2008 of 3 December 2008 implementing Directive 2007/2/EC of the European Parliament and of the Council as regards metadata, *Official Journal L 326, 04/12/2008 pp. 0012–0030* ("INSPIRE Metadata Regulation").
2. EC, Commission Regulation (EC) N° 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, *Official Journal L 323, 08/12/2010 pp. 11-102* ("INSPIRE Interoperability Regulation").
3. IFLA, ISBD(G): General International Standard Bibliographic Description, 2004 Revision, 39 pp., 2004. https://www.ifla.org/files/assets/cataloguing/isbd/isbd-g_2004.pdf/
4. MIG, Technical Guidance for INSPIRE Spatial Data Services and services allowing spatial data services to be invoked, Version 4.0, December 16, 2016. http://inspire.ec.europa.eu/documents/Spatial_Data_Services/Draft_TG_for_INSPIRE_SDS_2.0.1.pdf
5. EC, Directive 2007/2/EC of the European Parliament and of the Council of 14 March 2007 establishing an Infrastructure for Spatial Information in the European Community (INSPIRE), *Official Journal L 108, 25/04/2007 pp. 0001–0014* ("INSPIRE Directive").
6. De Rudder, A., An atmospheric metadata model, PASODOBLE D_CC-QUALITY_2.2/v1, August 26, 2011.
7. EC, INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119, EC JRC, 2013.
8. Christophe, Y., PAMIR internal progress report, May 18, 2015.

8. Web sites and downloadable documents

W1.	GeoNetwork	https://geonetwork-opensource.org/
W2.	ISO 19115-1/2014	https://www.iso.org/standard/53798.html
W3.	PAMIR	http://pamir.aeronomie.be/
W4.	BMDC	http://www.mumm.ac.be/datacentre/
W5.	OGC	http://www.opengeospatial.org/
W6.	ISO 19100 Geographic Information Standards	http://fght.utm.my/tlchoon/files/2016/02/ISO-19100-Geographic-Information-Standards.pdf
W7.	SeaDataNet	https://www.seadatanet.org/
W8.	Copernicus Marine Environment Monitoring Service	http://marine.copernicus.eu/
W9.	IASB-BIRA	http://www.aeronomie.be/
W10.	RMI	https://www.meteo.be/
W11.	NetCDF	https://www.unidata.ucar.edu/software/netcdf/
W12.	CF Conventions and Metadata	http://cfconventions.org/
W13.	IFLA	https://www.ifla.org/
W14.	QA4EO	http://qa4eo.org/

9. Acronyms

BMDC	Belgian Marine Data Centre [W4]
CDI	Common Data Index https://www.seadatanet.org/Metadata/CDI-Common-Data-Index
CF	Climate and Forecast [W12]
CoESRA	Collaborative Environment for Scholarly Research and Analysis http://www.tern.org.au/coesra-pg29647.html
CRS	Coordinate Reference System(s)
CSR	Certificate Signing Request
CTD	Conductivity-Temperature-Depth oceanographic sonde
EC	European Commission https://ec.europa.eu/
EDIOS	European Directory of the Ocean Observing Systems https://www.seadatanet.org/Standards/Metadata-formats/EDIOS
EDMED	European Directory of Marine Environmental Data https://www.seadatanet.org/Standards/Metadata-formats/EDMED
EDMERP	European Directory of Marine Environmental Research Projects https://www.seadatanet.org/Standards/Metadata-formats/EDMERP
EO	Earth Observation
EU	European Union http://europa.eu/
FP7	Seventh Framework Programme of the EU
GEOSS	Global Earth Observation System of Systems https://www.earthobservations.org/geoss.php
GMES	Global Monitoring for Environment and Security (previous name of Copernicus) http://www.copernicus.eu/

IASB-BIRA	Royal Belgian Institute for Space Aeronomy [W9]
IFLA	International Federation of Library Associations and Institutions [W13]
IMR	INSPIRE Metadata Regulation [1]
INSPIRE	Infrastructure for Spatial Information in the European Community https://inspire.ec.europa.eu/
IR	Implementing Rule(s)
ISO	International Organisation for Standardisation https://www.iso.org/
JRC	Joint Research Centre
LMO	Legally Mandated Organisation
MIG	INSPIRE Maintenance and Implementation Group
MIR	Metadata Implementing Rule
NetCDF	Network Common Data Form [W11]
NODC	National Oceanographic Data Center (USA) https://www.nodc.noaa.gov/
OGC	Open Geospatial Consortium [W5]
O&M	Observation and Measurement
PAMIR	Portal to Atmospheric and Marine Information Resources [W3]
PASODOBLE	Promote Air Quality Services Integrating Observations – Development Of Basic Localised Information for Europe (2010-2014)
QA4EO	Quality Assurance Framework for Earth Observation [W14]
RBINS	Royal Belgian Institute of Natural Sciences https://www.naturalsciences.be/
RMI	Royal Meteorological Institute [W10]
SDIC	Spatial Data Interest Community
SDS	Spatial Data Service

SQL	Structured Query Language https://fr.wikipedia.org/wiki/Structured_Query_Language
TWG	Thematic Working Group(s)
XML	Extensible Markup Language
WP	Work Package
WMO	World Meteorological Organization https://www.wmo.int/

Appendix A. INSPIRE mandatory metadata

INSPIRE mandatory metadata are listed in Table 1, which is a compact display of the contents of Parts B, C and D of the Annex to the IMR [1]. It is to be noted that “mandatory” has to be understood in a broad sense, namely that the provision of many metadata elements is conditional, for example, to the fact that the required information is defined, exists and is meaningful (conditions stated in Columns 5 and 6). Complementary information on the INSPIRE conventions and rules, as well as comments, are provided as end notes. Further detail on some metadata values are to be found in the IMR Annex.

Table 1. INSPIRE mandatory metadata.

Part B			***Part C***			***Parts B & D***		
#	**Name**	**Description**	**Multiplicity**[1]	**Definition restriction**	**Condition(s)**[2]	**Value domain**[3]		
						Num. ID (N)	**Textual name**[4] **(T)**	**Conventional name**[5] **(C)**
1. Identification								
1.1	Resource title	Characteristic and often unique name by which	1	/	/	/	*Free*	/

		the resource is known						
1.2	Resource abstract	Brief narrative summary of the content of the resource	1	/	/	/	*Free*	/
1.3	Resource type[6]	Type of resource being described by the metadata	1	/	/	1.1	Spatial data set series	series
						1.2	Spatial data set	dataset
						1.3	Spatial data service[7]	service[13]
1.4	Resource locator	Link(s) to the resource and/or to additional information	{0, 1, …}	/	If existing	/	/	*URL*
1.5	Unique resource identifier (URI)[8]	Value uniquely identifying the resource	{1, …}	/	/	/	/	*{code, namespace}*
1.6	Coupled resource	URI of the target spatial data set(s)	{0, 1, …}	RTCN = *service*	If existing	/	/	*{code, namespace}*
1.7	Resource	The language(s)	{0, 1, …}	RTCN =	If the resource	/	/	*3-letter character string as defined in the ISO*

	language	used within the resource		*dataset* or *series*	includes textual information Language(s) must belong to the ISO 639-2[9] list			*639-2 standard*[15]

2. Classification of spatial data and services

2.1	Topic category	High-level class(es) of the INSPIRE topic classification scheme – to assist in the topic-based search of available spatial data resources	{0, 1, …}	RTCN = *dataset* or *series*	/	2.1	Farming	farming
						2.2	Biota	biota
						2.3	Boundaries	boundaries
						2.4	Climatology / Meteorology / Atmosphere	climatologyMeteorologyAtmosphere
						2.5	Economy	economy
						2.6	Elevation	elevation
						2.7	Environment	environment
						2.8	Geoscientific	geoscientificInformation

							Information	
						2.9	Health	health
						2.10	Imagery / Base Maps / Earth Cover	imageryBaseMapsEarthCover
						2.11	Intelligence / Military	intelligenceMilitary
						2.12	Inland Waters	inlandWaters
						2.13	Location	location
						2.14	Oceans	oceans
						2.15	Planning / Cadastre	planningCadastre
						2.16	Society	society
						2.17	Structure	structure
						2.18	Transportation	transportation
						2.19	Utilities / Communication	utilitiesCommunication
2.2	Spatial data service type	Class of the INSPIRE service type classification	{0, 1}	RTCN = *service*	/	3.1	Discovery Service	discovery
						3.2	View Service	view

		scheme – to assist in the search of available services				3.3	Download Service	download
						3.4	Transformation Service	transformation
						3.5	Invoke Spatial Data Service	invoke
						3.6	Other Service	other

3. Keyword

3.1	Keyword	Word, formalised word or phrase commonly used to describe the resource subject – to help narrowing a full text search	{1, …}	/	/	If RTCN = *dataset* or *series* :		
						/	*Free but at least one concept from the GEMET thesaurus* [W14]	/
						If RTCN = *service :*		
						/	*Free but at least one concept from the IMR Annex* [R4], *Part D, Section 4. For example:*	/

						602	Temporal subsetting service	temporalSubsettingService
3.2	Originating controlled vocabulary	Citation of the controlled vocabulary from which the keyword value is borrowed (if borrowed from a controlled vocabulary)	{1, …}	For every keyword value provided	If relevant	/	*Citation including at least the title and reference date of the controlled vocabulary*	/

4. Geographic location

4.1	Geographic bounding box	Extent of the resource in the geographic space, given as a bounding box	{0, 1, …}	/	For RTCN = *dataset* or *series*, always mandatory For RTCN = *service*, mandatory if relevant	/	*Westbound and eastbound longitudes, southbound and northbound latitudes, in decimal degrees with a precision of at least two decimals*	/

5. Temporal reference

5.1	Temporal extent	Time period covered by the content of the resource	{0, 1}	/	If no value is given for 5.2, 5.3 & 5.4 [10]	/	*Individual date or interval of dates or mix of individual dates and intervals* *+ citation of temporal reference system used if not Gregorian & ISO 8601*[11]	/
5.2	Date of publication	Date(s) of publication or date of entry into force	{0, 1, …}	/	If no value is given for 5.1, 5.3 & 5.4 [16]	/	*Date(s)* *+ citation of temporal reference system used if not Gregorian & ISO 8601*[17]	/
5.3	Date of last revision	Date of last revision	{0, 1}	/	If relevant If no value is given for 5.1, 5.2 & 5.4 [16]	/	*Date + citation of temporal reference system used if not Gregorian & ISO 8601*[17]	/
5.4	Date of creation	Date of creation	{0, 1}	/	If no value is given for 5.1, 5.2 & 5.3 [16]	/	*Date + citation of temporal reference system used if not Gregorian & ISO 8601*[17]	/

6. Quality and validity								
6.1	Lineage	Process history and/or overall quality, optionally including 'Yes/No' statement on whether the data set has been validated or quality assured, whether it is the official version and whether it has legal validity	1	RTCN = *dataset* or *series*	/	/	*Free*	/
6.2	Spatial resolution[12]	Level of detail[13] of the data set	{0, 1, …}	/	For RTCN = *dataset* or *series*, if feasible For RTCN = *service*, if "a restriction exists on the spatial resolution"	/	*Set of zero to many resolution distances*[14] *or equivalent scales*[15]	/

7. Conformity

7.1	Specification	Citation of specification(s) to which the resource conforms[16]	{1, …}	/	/	/	*Citation(s) including at least the specification title and reference date*	/
7.2	Degree	Degree of conformity to the specification	{1, …}	For every specification provided	/	5.1	Conformant	conformant
						5.2	Not conformant	notConformant
						5.3	Not evaluated	notEvaluated

8. Constraint related to access and use

8.1	Conditions applying to access and use	Conditions for access and use including, where applicable, corresponding fees[17]	{1, …}	/	/	/	*Free.* *Alternatives:* · no conditions apply · conditions unknown	/
8.2	Limitations on public access	Limitations on public access imposed by an	{1, …}	/	/	/	*Free.* *Alternative:* no limitation on public	/

		EU Member State under Article 13 of the INSPIRE Directive, and reasons for them					access	

9. Organisation responsible for the establishment, management, maintenance and distribution of spatial data and services

9.1	Responsible party	Organisation(s)[18] responsible for the establishment, management, maintenance and distribution of the resource	{1, …}	/	/	/	*Organisation name*	*Organisation e-mail address*
9.2	Responsible party role	Role of responsible organisation	{1, …}	For every responsible party named	/	6.1	Resource provider	resourceProvider
						6.2	Custodian	custodian
						6.3	Owner	owner
						6.4	User	user
						6.5	Distributor	distributor

						6.6	Originator	originator
						6.7	Point of contact	pointOfContact
						6.8	Principal investigator	principalInvestigator
						6.9	Processor	processor
						6.10	Publisher	publisher
						6.11	Author	author

10. Metadata on metadata

10.1	Metadata point of contact	Organisation(s)[24] responsible for the creation and maintenance of the metadata	{1, …}	/	/	/	*Organisation name*	*Organisation e-mail address*
10.2	Metadata date	Date of creation or update of the metadata record	1	/	/	/	*Date conform to ISO 8601*[19]	/
10.3	Metadata language	Language in which the metadata are	1	/	/	/	*Name of an official language of the EC expressed in*	/

		expressed					*conformity with ISO 639-2*[15]	

Notes

[1] The multiplicity is the allowed number of occurrences of the metadata element characterising a single resource. INSPIRE has its own code to represent multiplicities, defined in Part C of the IMR Annex. Here, we indicate the set of the allowed multiplicity values, regardless of definition restrictions. Consequently, where some restriction applies, the set of multiplicity values will include "0". If, for example, the restriction is RTCN = *service*, the multiplicity value will indeed be 0 for RTCN = *dataset*.

[2] If no condition applies, the provision of the metadata element is mandatory where defined (when applying, definition restrictions are stipulated in Column 5). Otherwise, it is mandatory at the specified condition(s).

[3] Each value in the domain is a quadruplet {N, T, C, D} where N is the numerical identifier, T the textual name, C the conventional name and D an optional free-text description or definition of the value. Depending on the metadata element, some (but not all) of the components N, T, C and D can be empty.

[4] "Free" text which may be translated into a "natural language" different from English.

[5] Character string or set of character strings.

[6] The resource type conventional name will be abbreviated as RTCN in this table.

[7] In the IMR Annex, *T* = *Spatial data services* and *C* = *services* (plural). We suspect that this might be a mistake since anywhere else in the IMR a resource is said to be a *data set*, a *data set series* or a (= one) *service* (singular). To be investigated.

[8] This is a mistake made by the writers of the IMR. URI stands for "Uniform Resource Identifier" because its syntax is uniformly defined (as shown in the last column). The URI is not necessarily unique: one resource may have several URIs. But each of them points to a unique resource, so that it identifies it "uniquely" indeed.

[9] For sale at http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=4767 ; viewable (for example) at http://www.loc.gov/standards/iso639-2/php/code_list.php

[10] At least one of the metadata elements 5.1 to 5.4 must be provided.

[11] The default temporal reference system shall be the Gregorian calendar with dates expressed in accordance with ISO 8601 – cf. http://dotat.at/tmp/ISO_8601-2004_E.pdf

[12] This is actually the geographic resolution. N.B. No variable different than length is allowed. E.g., no angular distance is allowed (while angular distances are mandatory for the bounding box).

[13] This is actually the level of geographic detail of the data set.

[14] “A resolution distance is expressed by a numerical value associated with a unit of length”.

[15] “Equivalent scale is generally expressed as the (integer) value of the scale denominator”.

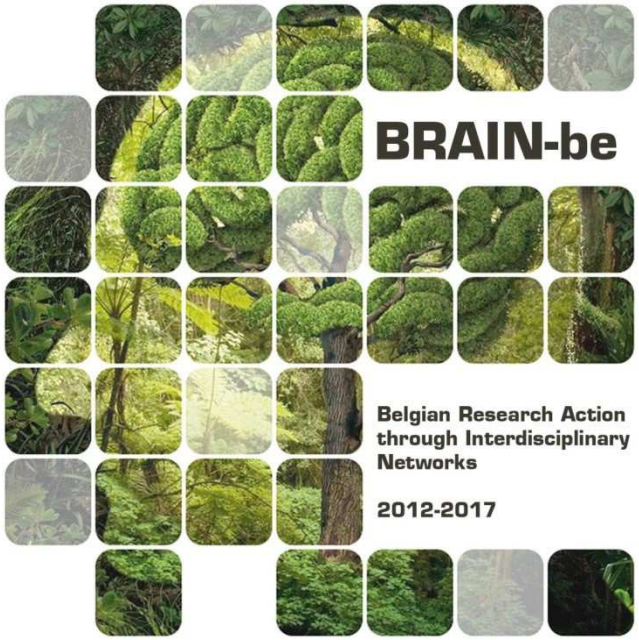
[16] Including, if applicable, implementing rules adopted under Article 7(1) of the INSPIRE Directive.

[17] As required by Articles 5(2)(b) & 11(2)(f) of the INSPIRE Directive.

[18] The IMR speaks of « the organisation ». Since, on the other hand, it specifies that the multiplicity of this metadata element is “at least one”, it is to be assumed that several organisations may be named (e.g. in the case of distributed roles).

[19] Cf. http://dotat.at/tmp/ISO_8601-2004_E.pdf

Appendix B. Metadata analysis



PAMIR Report on Metadata Appendix B: metadata analysis

T. Vandenberghe[1]

[1] Royal Belgian Institute of Natural Sciences (RBINS)

March 2018

PAMIR

A Portal to Atmospheric and Marine Information Resources
Brain-be Project BR/132/A6/PAMIR

1. Summary

Among the goals of PAMIR's 3rd Work Package on metadata was the identification of a set of common and field-specific features and the definition of a common metadata model. The oceanographic community, represented by BMDC, already makes use of the ISO 19115 standard for geospatial metadata in the CDI and EDMED community profiles. The ISO model is a good fit towards the needs of the oceanographic community. A metadata model capable of expressing the needs of the atmospheric community has been published in the PASODOBLE metadata standard proposal. To assess the fitness of the specific CDI profile and the generic ISO standard as a common metadata model, we present in this document an overview all the metadata fields of the PASODOBLE standard and propose mappings towards the ISO 19115 standard.

2. ISO 19115 and related geographical data standards

The original ISO 19115:2003 standard on geographic information - metadata dates from 2003. An XML schema implementation was published in three years later in 2006 in the ISO 19139 standard. A correction to ISO 19115:2003 was released at the same time. ISO 19115-2 was created in 2009 to accommodate imagery and gridded data. The most recent version, 19115-1, was published in 2014 and an XML implementation soon followed in May 2015. This implementation, 19115-3, is the successor of ISO 19139.

We list some of the changes between ISO 19115-1 and ISO 19115:

- Lineage moved from DataQuality to Metadata
- DataQuality specified in ISO 19157:2013
- Services metadata added from ISO 19119:2005
- Online linkage in Citation, dateStamp replaced by dateInfo,...

a. Extensions and community profiles of ISO 19139

ISO 19139 has many optional elements and can be extended. New metadata elements, entities, codelists, codelist entries, or more stringent obligations can be added. Some elements (i.e. mandatory) cannot be removed and make up the core components (i.e. the elements responsible for data discovery). The added elements are reported in MD_MetadataExtensionInformation, which itself is a part of the metadata. This is not modeled in ISO 19115:2003.

Removal of classes is best done in a profile, a community effort that canalizes more extensive changes into a new metadata model. In this profile, the community makes a select set of metadata elements mandatory and adds elements as needed.

Known ISO 19139 profiles are the SeaDataNet CDI profile, the World Meteorological Organisation (WMO) Core Profile, the Ocean Interoperability Platform (ODIP) Marine Community Profile, etc.

b. SeaDataNet CDI Metadata profile

The Common Data Index (CDI) metadata format is a marine profile of the ISO 19115-19139 spatial metadata content standard developed bij SeaDataNet. SeaDataNet is a standardized system for managing the large and diverse data sets collected by the oceanographic fleets and the automatic observation systems, and operates across European institutes. The latest version, 10, dates from september 2013. It provides an index to data on individual samples, cores and measurements and a unique interface to access these data sets online, the CDI Data Discovery and Access service. It covers physical, geological, chemical, biological and geophysical data, acquired in both European and global seas. The CDI format can be considered a de-facto standard for marine metadata in Europe. More than 1.7 million CDI entries are currently available. SeaDataNet has recently set-up a web service with virtual collections of CDI metadata entries. They are aggregated by discipline, data centre (based on controlled vocabularies), and geometric type (point/track/surface) which resulted in about 480 CDI collections. These virtual collections have no metadata attached except the metadata of the individual granules.

The XML encoding has been upgraded to the 19139 Schema, which is adopted as part of the EU INSPIRE Directive Implementing Rules. During this extension, the CDI schema has also been made INSPIRE-compliant.

Broadly, the index provides:

- The type of parameters it refers to
- The location and time of the measurement
- The method to retrieve the information
- Where applicable, basic preview (track)

During its history, it has been upgraded by Geo-Seas. Geo-Seas is a Pan-European infrastructure for the management of marine geological and geophysical data. Some of the improvements made available were:

- GML objects to support tracks and polygons next to points
- Service bindings for extra services (e.g. previewing)
- Resolution (spatial / temporal)
- Multiple instruments (to be able to include the positioning systems adjacent to the primary measuring device)

CDIs are created on a service/project basis: if a project of a service (i.e. research institute, data centre/NODC) has new data for which no CDI index exists yet, CDI indexes are created for each new sample in an automated fashion. For the discovery service, this has been done in a retroactive way. To access the data, either NetCDF files are available, or localized SQL views are created that contain all the data plus the CDI index. This index is a concatenation of service+project+sample. For each unique index in the view, dedicated software generates one CDI xml file that describes the metadata for this sample (because of this, the bounding box of one CDI metadata file is actually a point). Together with the CDI file, a mapping file points the column names coming from a controlled vocabulary to the local column name used in the view, and a coupling file registers the NetCDF file or the corresponding query from the SQL view.

c. SeaDataNet EDMED Metadata profile

The European Directory of Marine Environmental Data (EDMED) is a comprehensive reference to the marine data sets held within European research laboratories, data centres/NODCs. These datasets are the continuous data output of individual projects from their start to end. Just as CDI, EDMED has a metadata profile based on ISO 19115. CDI datasets are granules, the smallest data elements that make sense in an ocean science context. These CDI datasets can be aggregated to data series, which share or summate the common characteristics of the encompassed datasets and are themselves datasets. These data series might or might not be datasets in the EDMED context.

3. Metadata comparison between the PASODOBLE metadata model and the SDN CDI Profile

For any future developments, it is necessary to find a common ground of interoperability between PASODOBLE and CDI-ISO 19115. This common

ground could provide the foundation to any extensions or new metadata models of interest to the two science communities. The basis for this common ground is identifying the common elements. From this, the decision can be made to either develop a mapping scheme or to extend either the ISO 19115 model to a cross-community profile or try to nudge the CDI-ISO 19139 model into a common ground.

Given that the goal of this project is to explore a meaningful metadata interoperability between the European oceanographic and atmosphere science community and that currently, the CDI-ISO 19115 profile at a European level (in oceanography) and the ISO 19115 scheme or localized profiles at an international level are accepted spatial metadata standards, it is likely that the cross between the PASODOBLE metadata model and what the ocean community uses would make use of ISO 19115. The recent publication of ISO 19115-1 (2014) should be looked at in more detail.

As extending/profiling ISO 19115 could prove worthwhile, a crosswalk from ISO 19115 or CDI-ISO 19115 to PASODOBLE to identify common elements is less useful. It might still remain of interest to note the ISO 19115 entities and elements which the CDI-ISO 19115 did not adopt; however, upon extending, choices will have to be made on keeping these or not. The same elimination process would happen if one were to use a mapping.

A profile has little value on its own and is not worth the investment if it will not be used. An assessment of its necessity is needed.

On top of comparing the CDI implementation of ISO 19115, we took a first look whether the original, generic, ISO 19115:2003 edition provides better fitting elements for the functionality the PASODOBLE model envisages. For some fields, we made a comparison with ISO 19115:2014.

We list the steps taken to explore the value of ISO 19115 for the PASODOBLE model.

- perform a one-way crosswalk from PASODOBLE to CDI-ISO 19115 to have an idea of common (identical and similar) and noncommon elements.
- assess to what extent PASODOBLE can follow CDI-ISO 19115 and take a first look at how the non-common elements could be abstracted to ISO 19115 elements and entities.

a. Crosswalks

Crosswalks are the mapping of the elements, semantics and syntax between two (or more) metadata models or schemes. A scheme is composed of the element set and their encoding (syntax, most commonly in XML). Crosswalks can be either exploratory or result in a formal mapping (eg. expressed in

Xpath). Exploratory crosswalks result in a chart to identify the identical, similar or absent individual elements, their semantics and multiplicity. This is done just at the element level. Mappings are used to convert already existing metadata schemes from one to another, and are hence unidirectional.

There are two types of crosswalking. In absolute crosswalking, an absolute semantical equivalence is required. In relative crosswalking, semantical equivalence is preferred if possible but the mapping is more lenient. This ensures more metadata is preserved, but at the loss of exact meaning equivalence. When mapping it is possible to lose the information in a source compound element when the corresponding target element is 'atomic'.

b. PASODOBLE to CDI-ISO 19115/ISO 19115 crosswalk

In this crosswalk, the elements and their meaning have been compared between both models. As the PASODOBLE model itself is a proposal model, not achieving an absolute mapping is not a dealbreaker, so for this exercise, we will use a relative crosswalking whenever an absolute mapping is not possible.

The following commonality levels are distinguished:

- The elements are not common between PASODOBLE and ISO
- The elements are not common because the CDI metadata is only applicable to datasets and dataseries while the PASODOBLE scope is wider.
- The elements are relatively common: they are a syntactical close match (similar)
- The elements are (absolutely) common: they match syntactically precisely (identical)

Aside from commonality, the mapping should also make clear if an element would better map to more than one other element or vice versa. In most cases, a 1:1 map is possible, but some 1:n and n:1 element maps have been identified.

Lastly, there are elements that cannot be relatively or absolutely mapped, but can still be inferred from the presence of other mapped elements. The 'number of times' or 'has any x been performed' PASODOBLE elements fall under this.

The crosswalk contains the following elements:

- *cdi element node path and cardinality per node (if not 0 or 1)*

- *final cdi nodes and cardinality with respect to parent node*
- *mapping cardinality: 1:n mappings contain n final cdi nodes*
- *cdi element definition*
- *remarks*

We have listed all the PASODOBLE elements in an tabular crosswalk overview and have mapped them with ISO elements. In the further discussion (see below) we only discuss the PASODOBLE elements for which no identical field match in ISO has been found. To give an idea, of the 115 PASODOBLE fields, 59 have no direct mapping in CDI-ISO 19115, 35 of these can be expressed in ISO 19115:2003 and 20 can't be expressed in ISO 19115:2003 (17%). The results of the textual mapping (this document) has been complemented in the tabular crosswalk.

c. General PASODOBLE-CDI-ISO 19115 differences

Both metadata models differ in objective and scope.

Variables

The CDI-ISO 19115 model can describe any number of variables (dependent physical quantities) present in a dataset. The PASODOBLE proposal specification has been written in view of one variable per metadata instance. The drawback of this is that for each variable in a dataset another metadata instance is needed. The PASODOBLE model leaves open the possibility to extend it to any number of variables.

The CDI-ISO 19115 scheme provides the possibility to report some of the independent variables of a dataset through the axes. However, this is not optimal.

In the PASODOBLE scheme, it is possible to define validation methods and quality indicators for each dependent variable, which is not possible in CDI-ISO 19115.

Data series/data sets

The CDI service exposes data sets at the granule level, which can be aggregated in any way into data series. In ISO 19115 these aggregated datasets are referrable to (='citable') just as the metadated dataset can be referred to (='cited'). In the CDI-ISO 19115 scheme the identifier of a cited dataset can be the EDMED identifier, and it is mandatory to provide one (exactly one) type of aggregation initiative. This is interesting as it provides a mechanism on how the parent dataset has been broken down (or built up) and ultimately what the granules are all about. Examples are: campaign, collection, experiment, mission, platform, process, project,...

In practice both aggregation and disaggregation are possible; metadata-wise, these are represented in the same way bottom-up way.

In the CDI-ISO 19115 scheme there are many ways to refer to the parent of a dataset, a dataset series:

- by having the parentIdentifier (in MD_Metadata) refer to the fileIdentifier (unique identifiers for actual files) of another dataset (which should be a series).
- by circumscribing the series (via three FreeText fields) in the self-citation of the dataset. The series could still be described in a separate file.
- by using the aggregationInfo element of identificationInfo, then either
  - citing the aggregated dataset (which in turn could name its own aggregate dataset)
  - identifying the aggregated dataset, in which case the EDMED code can be used

Using aggregationInfo has the benefit of being able to use the aggregation initiative (why has the dataset been aggregated).

In CDI-ISO 19115 it is not possible to have one metadata instance describe multiple datasets (i.e. have multiple instances of identificationInfo), however it is in ISO 19115:2003. This would provide a way to describe a data series and all its (possible, i.e. note that datasets can be split up dynamically) subsets in one metadata instance (i.e. file), but this would be very cumbersome. Even more so as the bidirectionality would need to be internally consistent. This is typically something to restrict in a profile.

The PASODOBLE model doesn't have an aggregation model to express that a dataset is a subset of another, but it has strong set of elements describing the product derivation, a concept that is present in the ISO 19115 LI_Lineage class.

Data product/data service

The PASODOBLE metadata schema is designed to jointly provide metadata to coupled information resources, be them data sets, data products, data services or data models. The metadata element set pertaining to the data product is an aggregation of (any) metadata elements pertaining to the data service(s). This is in contrast to INSPIRE where there is an association between the metadata sets for products and services: the service can have any coupled resources on the condition that they are available via the service. The two INSPIRE data service/dataset-data series sets of elements are just slightly different. Of note however is that the association between both is unidirectional: dataset metadata should not contain a description of the service(s) it is made available in.

Nonetheless, the PASODOBLE model proposes to keep the linkage between service and product of both aggregation and association type, so that services can exist on their own, i.e. can have metadata describing them without having any (reported) data product (data sets without data service are also possible in PASODOBLE). With this in mind, the simpler INSPIRE-like association model would not necessarily be a bad fit (albeit made bidirectional).

As the data circumscribed by CDI metadata only has meaning within the context of the SeaDataNet Common Data Index Data Discovery and Access service, the data product is always tied to the same service. The need of composing the data product from the data service(s) is less useful in this case. The service is not a separate entity in ISO 19115:2003, but it is circumscribed in the distributionInfo element of MD_Metadata.

ISO 19119 is a separate standard for service implementations and provided the service metadata elements that are reused in ISO 19115-1, via the MD_ServiceIdentification class. This class reflects the structure proposed in PASODOBLE with regards to the integration of the data service inside the data product metadata, and at the same time keeping it possible to metadate a service.

The CDI-ISO 19115 marine profile can only describe metadata of data sets (or data series), as both its /hierarchyLevel as its /hierarchyLevelName make use of a restricted MD_ScopeCode class and CodeList (resp. 'dataset', 'series' and 'Common Data Index Record'). ISO 19115 provides a plethora of scopecodes (MX_ScopeCode), though, including 'model', 'service', 'software', 'nonGeographicDataset' etc. This is a PASODOBLE requirement.

Dynamic data

The PASODOBLE History and Maintenance and update sections acknowledge the existence of changing data by providing the user with the frequency

with which data is reassessed, but can only describe the current version and the changes between the previous one. CDI-ISO 19115 provides a similar set of elements.

Providing a truer representation of the history of the dataset is beyond the scope of this exploration. In general, there are two approaches to describe history: snapshots vs. changesets.

d. Crosswalk of individual PASODOBLE elements

The PASODOBLE (sub)section elements have been mapped to the elements of CDI-ISO 19115. For each section the non-common elements and their wider scope are discussed.

A. Product

4. A.1 Identification and nature

- **Product type**: The element product type is present in CDI-ISO 19115 but is restricted to 'dataset' or 'series'. ISO 19115 allows many more product types.
- **URI of the target data product(s)**: Meant is the URL if the product is a tool intended to support some specific data product(s). The metadata of the CDI-ISO 19115 scheme can only describe datasets. ISO 19115:2003 can add the service url info via the SV_ServiceIdentification class and its onlineResource.
- **Nature of dependent variable**: The PASODOBLE model proposal is designed for one dependent variable to be reported but proposes to extend this to multiple variables (this would otherwise lead to one metadata file per variable). CDI datasets commonly contain more than one parameter, but these all have the same origin. Extending PASODOBLE in this respect entails that all of the fields in A.1, A.2, A.3 and A.4 would have cardinalities higher than 1; it is possible and even probable that information common to multiple variables would be shared (for example bounding box).
    - Furthermore, the PASODOBLE model proposes to allow metadata to describe both datasets of variables measured at the same time, and of variables that are stored in the same file but aren't necessarily measured at the same time, can have different origins, might have

different bounding boxes, have different independent variables etc. As parameters in CDI-ISO 19115 are expressed simply as instances of MD_Keywords (with a controlled vocabulary), it is impossible to convey this information. The element mapping rests on the assumption that all elements in sections A.1, A.2, A.3 and A.4 apply to just one variable or that they are identical for any variables. In ISO 19115 this is possible due to the unlimited multiplicity of the identificationInfo element.

- ISO19115 has two separate classes for describing the data parameters of gridded data (continuously varying data modeled, or calculated from points and represented in a grid); and those of vector and point data (usually discrete data represented as points, lines and polygons) in the Content Information package. These are resp. MD_CoverageDescription for describing several types of grids and MD_FeatureCatalogDescription, both subclasses of the abstract MD_ContentInformation class. This is a relatively cumbersome class. The Marine Community Profile[7] developed by the Australian Ocean Data Centre and ODIP simplifies this by making a dataParameters element in MD_DataIdentification available (instance of DP_DataParameter). Each parameter has one or more names, units, a minimum and maximum value and a description. This coincides with parts of the PASODOBLE proposal. Absent are the multiplicities of A1. to A4., especially the validation.

- **Unit:** the unit of any dependent variable is not present in CDI-ISO 19115 (it is given in the dataset itself). It can be expressed in ISO 19115:2014 using the MD_SampleDimension/units element.
- **Type of data sourcing**: what the global type of the ultimate data sources is, is not present in CDI-ISO 19115. The ISO 19115 LI_Lineage class allows to document all sources and the process steps needed to transform them, but there is no direct way of assigning a general type to the first source in the chain.
- **Immediacy**: the actuality (eg. being an observational and close-ended data record, a near-real-time record or a forecast) a of the record is not present in CDI-ISO 19115.
- **Spatial representativeness**: an indication of how the record as stored is representative for the area it is measured in (point data, summed or averaged) is not present in CDI-ISO 19115. It can be expressed in ISO 19115:2003 in the MD_SpatialRepresentation/axisDimensionsProperties/MD_Dimension class. However, this allows only a numeric expression of the degree of

7 http://mcp-profile-docs.readthedocs.io/en/stable/extensions/dataParameters/index.html

detail in the dataset. In ISO 19115:2014, this class has been extended to not only include the dimension name, size and resolution, but also a textual description of the dimension, in which averaging or summation of the values along the axis (in this case x, y or z) can be indicated with free text. Specifically the ISO 19115:2014 Coverage classes could be of interest to further evaluate the mapping of this field and the following, specifically MD_SampleDimension.

- **Temporal representativeness**: an indication of how the record as stored is representative for the temporal range it is measured in is not present in CDI-ISO 19115. A similar technique as with the spatial reprentativeness can be used, but then the value 'time' for the MD_DimensionNameTypeCode of the MD_Dimension class instance can be chosen.
- **Value representativeness**: an indication of how the record value as stored is representative for the actual in situ value is not present in CDI-ISO 19115. A similar technique as with the spatial reprentativeness can be used, but a more detailed analysis of the values for the MD_DimensionNameTypeCode would be needed.

5. A.2 Domain of definition

- **Spatiotemporality**: an indication whether the independent variables are spatiotemporal or not is not literally present in CDI-ISO 19115. It can be approximated by using MD_SpatialRepresentation/axisDimensionsProperties/MD_Dimension and by analysing the occurrence of the MD_DimensionNameTypeCode values x, y, z, and time.
- **Number of independent variables**: The independent variables are recorded in the MD_SpatialRepresentation/axisDimensionsProperties/MD_Dimension. A count of the number of instances would give a mapping to this field.
- **Independent variables**: the independent variables are represented by the axes in the CDI-ISO 19115 scheme. Axes can only provide partial information on the independent variables as they can't include any variables other than x, y, z, track, line, sample, time, etc. (possible MD_DimensionNameTypeCode) and can be of a compound type (tracks and lines).
  - **Unit**: the unit of the independent variable is recorded in CDI-ISO 19115 scheme as the Measure (unit + quantity) of the axis resolution, but this is not necessarily the one used for the independent variable itself.
  - **Boundaries**: the boundaries in which the variables are applicable are indirectly represented in the CDI-ISO 19115 Vertical, horizontal

and temporal extent entities.

6. A.3 Domain of values

- **Confidence level**: The probability that a data value belongs to the confidence interval is not present in CDI-ISO 19115. It has been also found to be absent in ISO 19115:2003.
- **Confidence interval**: This is not present in CDI-ISO 19115. It has been also found to be absent in ISO 19115:2003.

7. A.4 Quality

In PASODOBLE, each dependent variable can be validated by testing for any number of quality indicators. In CDI-ISO 19115, as in ISO 19115 validation does not operate on the different variables (which are keywords), but on itself. In PASODOBLE, quality indicators can be of two categories: those that differ from data point to data point or those that characterise the whole dataset. Quality in CDI-ISO 19115 is assessed in terms of conformance (ConformanceResult) of result to a standard specification. This is more alike to the PASODOBLE concept put forward in A.7 References. The simplified CDI-ISO 19115 profile omits true Quality indicators and only uses a text-based derivation Lineage entity and the ConformanceResults entity (both part of the DQ_DataQuality class).

The ISO 19115 DQ_DataQuality class can express most of the ideas put forward in the PASODOBLE model, save the link with the actual variables and the difference between individual and global QIs. The (abstract) DQ_Element class contains elements as nameOfMeasure, measureDescription, evaluationProcedure, etc. It is possible to apply a scope to each DataQuality instance, but although the scope code list is long, individual data points are not addressable. In contrast to CDI-ISO 19115 it is also possible to have quantitative test results (DQ_QuantitativeResult). Confidence intervals and levels can be entered into the errorStatistic and value text elements. DQ_Abstract has a number of nonabstract subclasses that are provided to specify the type of reports: e.g. DQ_ConceptualConsistency, DQ_PositionalAccuracy, DQ_QuantitativeAttributeAccuracy,...

8. A.5 History

The history of the dataset (date of creation, version number, version issue date) is described in the IdentificationInfo self-reference (Citation).

- **Change between current and previous versions**: This is implicitly included in CDI-ISO 19115 in the Lineage element, not as a step

description, but as free text. In full ISO 19115:2003, the LI_ProcessStep element allows identifying each separate step in the dataset history, complete with a date, rationale and responsible party.

9. A.6 Presentation, coding, format

- **Display**: Being a sort of classification for the computer format, this is not included explicitly in the ISO 19115:2003 model. The MD_Identification/citation/CI_Citation/presentationForm/CI_PresentationFormCode codelist provides a similar classification scheme. Examples from the codelist are documentDigital, imageDigital, profileDigital,...
- **Computer language**: The programming language of the model, service etc. is not included in CDI-ISO 19115 as it only deals with datasets. Similar to this, in ISO 19115, the application schema used to build the dataset (i.e. tables, columns, relations, and constraints) is included in the MD_ApplicationSchemaInformation (schemaLanguage, constraintLanguage, softwareDevelopmentFile,...).
- **Computer format(s)**: The data format of the data product is provided in the distributionFormat element (of multiplicity 1...n) of MD_Distribution, which is an instance of MD_Format. The version of the format is also included. We opted for a relative mapping as the delivery service does not necessarily operate on discrete files, causing the data service output format type and the original to be incomparable. The CDI service system can map the data from each NODC database to an autogenerated odv file, but it can also operate on NetCDF files.

10. A.7 References

- The standards applicable to the data product, its specifications or its validation can be referenced in CDI-ISO 19115 as DataQuality reports, each with a scope (in CDI limited to either the dataset itself or the data series it is part of). ConformanceResults can't be used for backreferences (i.e. the dataset would then be validated by its own description, which is always a pass).
    - However, it is not possible to limit the scope to certain moments in the data lifecycle (i.e. a report with initial dataset development) in CDI-ISO 19115. The specification element (which is an instance of CI_Citation) of the DQ_ConformanceResult cites the standard in question. In ISO 19115 it is possible to specify the dateTime of a DQ_Element instance (i.e. a DataQuality report), which in turn dates all ConformanceResult and QuantitativeResult instances the DQ_Element has. As it is possible to also date each processStep in the Lineage entity, this would provide a mechanism to model the data life cycle 'scope' proposed in the PASODOBLE model.

- The data product as a whole, the product specifications or the product validation; and the product development steps are definitely modellable in the LI_lineage class. However, the services associated to the product fall out of this scope; note also that LI_lineage and DQ_Element classes are part of the DQ_DataQuality class. CDI-ISO 19115 is not fit to describe service information, ISO 19115 is (with restrictions on references). See C. Product availability, distribution and usage.

- **Degree of conformity to the standard**: this degree is expressed in CDI-ISO 19115 as a boolean pass/not pass element within DQ_ConformanceResult (has the dataset passed the cited specification?). In CDI-ISO 19115, it's not possible to refine the degree; in ISO 19115 this is a boolean as well.
- **Publication(s)**: publications describing, using or referring back to the data product or its validation can be absolutely mapped to CDI-ISO 19115 SDN_DataIdentification/additionalDocumentation. In ISO 19157/19115-1 it is possible to use DQ_StandaloneReportInformation to report data quality information outside of the dataset.

B. Product generation

11. B.1 Developer(s)

The developers, persons or institutes responsible for the initial creation and data gathering of the dataset and for its initial data management, which includes the Principal Investigator, are included in CDI-ISO 19115 and in ISO 19115 as one of the responsibleParty instances of the self-citation of the resource. The general contact persons for the dataset (i.e. MD_Identification/pointOfContact), is in a CDI context usually a National Oceanographic Data Centre (NODC). ISO uses a role-based system for the ResponsibleParty class, with which finer responsibilities can be expressed. The roles available in CDI-ISO 19115 are resourceProvider, custodian, owner, user, distributor, originator, pointOfContact, principalInvestigator, processor, publisher and author. It is clear that the PASODOBLE terms developer, provider and support can map to multiple ISO roles. For this comparison we map 'developer' to 'originator', 'provider' to 'distributor' and 'support' to 'pointOfContact'. 'Principal investigator' maps absolutely.

The project or programme that led to the initial creation of the dataset can be included in CDI-ISO 19115 via any descriptiveKeywords keywords. It is advisable to group multiple related keywords together in one MD_Keywords instance as the type (e.g. 'project') can be specified as well.

With these caveats, most elements are mappable to CDI-ISO 19115 and to ISO 19115.

- Project principal investigator(s) (PI): The PASODOBLE model proposes to record for each project the Principal Investigator. We propose a relative mapping to CDI-ISO 19119 as it is easy to list the PIs separately in the citedResponsibleParty entity, but providing the PI for each project is impossible, as generally, in ISO 19115 the project is just a keyword of type project. In ISO 19115 this could be approximated by the positionName element (CharacterString) of the CI_ResponsibleParty class.

12. B.2 Derivation chain and intermediate validation

The lineage in CDI-ISO 19115 is just a string expressing information about either events or source data used in constructing the data specified by the scope, or lack of knowledge about the lineage (the statement). SeaDataNet hasn't issued any rules on how to delimit stages in the string. The boilerplate sentence that the SeaDataNet metadata editor provides ('*The data centres apply standard data quality control procedures on all data that the centres manage. Ask the data centre for details.*') is relatively uninformative; the statement attribute was added in order to be INSPIRE compliant. The summary text of all these stages is relatively mappable to the CDI-ISO 19115 statement (n to 1).

This is quite a simplification compared to the possibilities in ISO 19115 itself. There several Source instances plus several ProcessStep instances can be part of the the Lineage entity. These Sources represent stages. Each Source is characterized by a name (via a citation), a description, an SRS, and an extent (vertical, horizontal and temporal). The ProcessStep entity has a description, rationale, time and processor. Algorithms and validation (as per the PASODOBLE model proposal) are not as such present but could be represented in the ProcessStep description.

13. B.3 References

The PASODOBLE model proposes to provide a total overview of references for the steps in the derivation chain. As described in A.7 References, this is possible via DQ_ConformanceResult and additionalDocumentation. In CDI-ISO 19115 it is not possible to assign a scope or date to a ConformanceResult instance, though in ISO 19115 it is possible to at least assign a date.

- Publication(s): reference(s) of publication(s) describing the product development can be relatively mapped to CDI-ISO 19115 SDN_DataIdentification/additionalDocumentation. Note that there is a distinction between the dataset itself and the validation steps that operated on it. Most likely the SDN developers mean that the subject of this element is 'the mature dataset', so the dataset plus the validation steps. In ISO 19157/19115-1 it is possible to use DQ_StandaloneReportInformation to report data quality information outside of the dataset.

C. Product availability, distribution and usage

ISO 19115 includes the SV_ServiceIdentification class, which together with MD_DataIdentification inherits from the abstract class MD_IdentificationInformation (meaning all elements in MD_IdentificationInformation are present in both child classes). MD_Metadata/identificationInfo[1...n] can be an instance of either MD_DataIdentification or SV_ServiceIdentification.

14. C.1 Provider

This section contains information about the service and the provider of the service.

As described in B. 1 Developers, we map 'provider' to the ISO 19115 CI_RoleCode 'distributor'. In the SeaDataNet CDI context, the provider is SeaDataNet itself. The same remarks as in B.1 Developer(s), save the one about projects, apply.

- **Service name**: this is not included in CDI-ISO 19115. In ISO 19115, the service name is represented in the SV_ServiceIdentification/citation/CI_Citation/title element.
- **Provider acronym**: this is not included in CDI-ISO 19115 nor in ISO 19115.
- **Service operation start date**: this is not included in CDI-ISO 19115. In ISO 19115 it is possible to use SV_ServiceIdentification/status to express the 'freshness' of the service. This makes use of the MD_ProgressCode codelist (completed, historicalArchive, obsolete,...).

15. C.2 Applications

The application section lists the purpose of the whole service, the purpose in terms of its applications ('Targeted application(s)'), and contains a number of examples of the applications: what is the name of the application, who/what is using ('applies') the application. Applications are understood as general functionality, not dedicated subservices.

This whole section is absent in CDI-ISO 19115 and is present in limited form in the elements of the SV_ServiceIdentification class of ISO 19115.

- **service purpose**: present in ISO 19115 as SV_ServiceIdentification/purpose
- **Targeted application(s)**: It is difficult to assess whether 'applications' can be modelled in ISO 19115. The

SV_ServiceIdentification/containsOperations element (instance of SV_OperationMetadata) might be suited, but is rather intended for dedicated subservices.

- **Number of example(s)**: not present in ISO 19115.
- Examples:
  - **description**: a description of an application of the service
  - **user name/user acronym/e-mail address**: characteristics of the institution or individual that applies the service in the above defined way: not present in ISO 19115.

16. C.3 Maintenance and update

PASODOBLE separates between data updates and data revisions/reassessments and considers them to be service-specific. Data updates comprise new data, ingested into the dataset, while data revisions or reassessments comprise changes made to existing data ('maintenance'). The changes can each have a different volume ('revision', 'reassessment') and can be organised in whatever way (eg. iterative, ad hoc).

In ISO 19115 resourceMaintenance is an element of MD_IdentificationInformation, meaning that resourceMaintenance can both be expressed at a dataset level as at a service level. In CDI-ISO 19115 there is no separation between updates and revisions: the maintenanceAndUpdateFrequency can list both. It is conceivable that this could nonetheless be refined by using the userDefinedMaintenanceFrequency to express pure maintenance ('revisions') but there is no guideline for this from SeaDataNet.

In ISO 19115 it is possible to (beyond the scopes in MD_ScopeCode) circumscribe the scope of a MaintenanceInformation instance by using its updateScopeDescription (instance of MD_ScopeDescription) element. With this, the scopes 'update' or 'maintenance' could be defined.

There is no possibility in CDI-ISO 19115 nor ISO 19115 to indicate data backupping and archival literally in the fine-grained way PASODOBLE proposes. It is possible to use 'repository' as an updateScope in the MD_MaintenanceInformation class in ISO 19115:2003.

- **Update frequency**: This is included in CDI-ISO 19115 in the MD_MaintenanceInformation fclass as a vocabulary value (maintenanceAndUpdateFrequency) or as a timeperiod (userDefinedMaintenanceFrequency).

- **Update delay**: This is not included in CDI-ISO 19115. It can be conveyed in ISO 19115:2003 by providing a date in dateOfNextUpdate.
- **Backup frequency**: This is not included in CDI-ISO 19115 nor ISO 19115. It could be conveyed as a maintenanceNote.
- **Archive maintenance**: This is not included in CDI-ISO 19115. It is possible to indicate in ISO 19115:2003 by using an additional MD_MaintenanceInformation instance and setting 'repository' as an updateScope and detailing how the repository is maintained.
- **Archive coverage**: The time period covered for a specific archive is not included in CDI-ISO 19115 nor ISO 19115.
- **Archive availability**: Whether the archive is partially or totally available is not included in CDI-ISO 19115, but can be expressed textually in MD_MaintenanceInformation/maintenanceNote.
- **Available archive time coverage**: This is not included in CDI-ISO 19115 nor ISO 19115.

17. C.4 Delivery

A data product could be transferred in several different ways. In CDI-ISO 19115 all possible delivery ways of the data product are described in the transferOptions element of the MD_Distribution class. Both a transferSize (in Mb) and the different onLine resources can be further specified, each with url, protocol, name, description and general function type.

- **Delivery mode(s)**: The delivery mode maps (relatively) to the CI_OnLineFunctionCode codelist. This codelist lists the functions the service is capable of (download, information, offlineAccess,...). These codes encompass both the way in which the service is delivered as their terms & conditions, therefore also the PASODOBLE Terms and conditions/Access conditions element maps to this ISO element.
- **Delivery channel(s) or medium(s)**: This element can be mapped (relatively) to protocol.
- **Delivery frequency**: The periodicity or period with which the product is delivered is not included in CDI-ISO 19115. A related concept is present in ISO 19115 as MD_StandardOrderProcess/plannedAvailableDateTime.
- **Delivery delay**: This is not included in CDI-ISO 19115. It is present in ISO 19115 as MD_StandardOrderProcess/turnaround.

18. C.5 Terms and conditions

- **Usage conditions or restrictions**: in CDI-ISO 19115 these conditions are described in the MD_LegalConstraints class. The 0 to n accessConstraints are restricted by a codelist. A relative mapping is maximized if any usage conditions or restrictions not present in the MD_RestrictionCode are represented in the otherConstraints. ISO 19115 adds the useConstraints which also make use of the MD_RestrictionCode codelist.
- **Access conditions**: The access conditions maps (relatively) to the CI_OnLineFunctionCode codelist. This codelist lists the functions the service is capable of (download, information, offlineAccess, downloadRegistration...). downloadRegistration e.g. means '*manual interaction with an on-line system by registered users following successful authentication and authorisation*'. In CDI-ISO 19115 there is no mention of a fee (the CDI service is free). These codes encompass both the way in which the service is delivered as their terms & conditions, therefore also the PASODOBLE delivery mode maps element maps to this ISO element. In ISO 19115 it is possible to describe the standard order process (MD_StandardOrderProcess). One can give details on fees, date and time when the resource is available, general instructions on getting the resource and the typical turnaround time for an order.

19. C.6 Support

Support is not as such present in CDI-ISO 19115, but the MD_Identification class pointOfContact element could denote a support function. Only the PASODOBLE elements pertaining to a citedResponsibleParty (eg. Help desk telephone/email) are expressible in CDI-ISO 19115. Note that the support isn't necessarily performed by the distributor.

ISO 19115 has the MD_Usage class (which has been left out of CDI-ISO 19115). The MD_Identification class can have 0...n usage instances. This entity provides a manual rather than a description of the support. In any case, the MD_Usage class userContactInfo element maps absolutely to the PASODOBLE Help desk telephone/email elements.

- **Language(s) used to provide support**: This is not included in CDI-ISO 19115 nor ISO 19115.
- **Additional service(s) provided**: This is not included in CDI-ISO 19115. ISO 19115 provides the SV_ServiceIdentification/containsOperations (instance of SV_OperationMetadata class) which can express the operations a service can perform. Definition: '*Description of one's service*

operation. Together, attributes and components provide the signature of the operation (i.e. the method).'

- **Alternative formats that can be automatically generated**: all formats available in the service are included in CDI-ISO 19115 in the distributionFormat (1...n instances of MD_Format) of the MD_Distribution class. ISO 19115 can discern between the data format of the dataset, reported in the distribution info of the dataset (i.e. MD_Metadata/distributionInfo/MD_Distribution/distributionFormat/MD_Format), the resource format of the dataset itself (i.e. MD_Metadata/identificationInfo/MD_DataIdentification/resourceFormat/MD_Format) and the data format in which the data is served (i.e. MD_Metadata/identificationInfo/SV_ServiceIdentification/resourceFormat/MD_Format). This mechanism could help to discern between the primary data format served and the alternative served formats, through a conversion service. The SDN profile implementors have represented the data format as served in the distributionFormat.
- **Data handling options provided**: This is not included in CDI-ISO 19115. In ISO 19115 this could be modelled in SV_ServiceIdentification/containsOperations.
- **Other**: This is not included in CDI-ISO 19115. In ISO 19115 this could be modelled in SV_ServiceIdentification/containsOperations.

20. C.7 References

The PASODOBLE model proposes to provide references to the standard(s) applicable to the development of the services associated to the product. ISO 19115 DQ_ConformanceResult, CDI-ISO 19115 additionalDocumentation, ISO 19115-1/19157 DQ_StandaloneReportInformation are unfit to describe the whole data product lifecycle from the creation of the data product to the implementation of the distribution and delivery of the data product. It is currently unclear how this could be solved in ISO 19115.

- **Standard(s) of reference**: This is not included in CDI-ISO 19115.
- **Publication(s)**: This is not included in CDI-ISO 19115.

D. Metadata

Information about the metadata itself is well-covered in CDI-ISO 19115. First we can consider the main metadata standard which is described in MD_Metadata/MetadataStandardName + metadataStandardVersion. However, the PASODOBLE model proposes it should be possible to refer to

multiple metadata standards. In CDI-ISO 19115 this is achieved by using the DQ_ConformanceResult class. The designers of CDI themselves use this path to express the adherence of the CDI model to INSPIRE (aside from ISO 19115). However, strictly speaking, the DQ_ConformanceResult is a part of DataQuality, which pertains to the data itself, not to the metadata standard.

- **Date of last update**: A date of revision is not present in CDI-ISO 19115 nor ISO 19115 but can be represented in ISO 19115-1. In the new edition (2014), MD_Metadata can have multiple CI_Date instances instead of just the dateStamp. The type MD_DateTypeCode (to be used in the CI_Date instance) has been extended and now has 'lastUpdate'.
- **Standard(s) of reference**: In CDI-ISO 19115 the name (and version) of the main standard are expressed differently as the name (and version) of any additional standards (see higher). This same approach is advisable in ISO 19115 as its multiplicity is 1 as well.
- **Degree of conformity to the standard**: The pass result is only possible for the additional standards. One can argue that this is only meaningful for the additional standards given that the main metadata standard is fulfilled by running the metadata file in a metadata validator.

Metadata crosswalk

Legend: not common | not common, as CDI only deals with datasets | Common-identical (absolute match) | Common-similar (relative match) | Not common. Can be derived from presence/absence/counting

paso_index	paso_par	paso_title	paso_field	paso_desc	inspire number	paso_remark	cdi element path/cardinality	multiplicity	mapping multiplicity	Not common?	Resolved in ISO 19115:2003	cdi element name	cdi definition	remarks
1	A.1	Identification and nature	Product name	/	1.1		identificationInfo/SDN_DataIdentification/citation/CI_Citation/	1	1:1			title	name by which the cited resource is known	
2	A.1	Identification and nature	Product type	[value] = 'Data' / 'Service' / 'Other'	1.3					NC	y	not common, as CDI only deals with datasets		Both /hierarchyLevel and /hierarchyLevelName make use of a restricted MD_ScopeCode class and CodeList (resp. "dataset", "series" and "Common

														Data Index Record"
3	A.1	Identification and nature	Definition / Description	Summary of the product content. When the product is data-based and the data not a commonly used physical quantity, a rigorous definition or description of how the quantity is computed (what its components are and how they are combined) and /or what it represents should be provided. If the product is a model, description of the model including, when relevant, modelled physical processes, parameterisations, boundary and initial conditions, diagnostic variables.	1.2		identificationInfo/SDN_DataIdentification/	1	1:1			abstract	brief narrative summary of the content of the resource(s)	
4	A.1	Identification and nature	Keyword(s)	/	3.1		identificationInfo/SDN_DataIdentification/descriptiveKeywords[1...n]/MD_Keywords/	1...n	1:1			keyword	commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject	
5	A.1	Identification and nature	Standard terminology	For each keyword provided, if the keyword is borrowed from a vocabulary standard, citation of this standard	3.2		identificationInfo/SDN_DataIdentification/descriptiveKeywords[1...n]/MD_Keywords/	1	1:1			thesaurusName	name of the formally registered thesaurus or a similar authoritative source of keywords	
6	A.1	Identification and nature	INSPIRE category	INSPIRE topic category or categories (for a data product) or (unique) service type (for a non-data product)	2.1 + 2.2		identificationInfo/SDN_DataIdentification/topicCategory	1	1:1			MD_TopicCategoryCode	main theme(s) of the resource	This codelist lists all the GEMET categories. Restricted to 'oceans' for CDI.
7	A.1	Identification and nature	URI of the target data product(s)	If the product is a tool intended to support some specific data product(s)	1.6	Only applicable when the product is a tool intended to support data products				NC	y	not common, as CDI only deals with datasets.		Not possible in CDI. The service that distributes the data itself is part of the scheme (DistributionInfo).

8	A.1	Identification and nature	Nature of dependent variable				identificationInfo/SDN_DataIdentification/descriptiveKeywords[1...n]/MD_Keywords/keyword/	1...n	1:1			SDN_ParameterDiscoveryCode	commonly used word(s) or formalised word(s) or phrase(s) used to describe the subject	This codelist lists all possible parameters in oceanography. Parameter and platform class are the two mandatory keyword elements. Note that the PASODOBLE model describes one set of metadata per dependent variable in a dataset. The 1:1 mapping cardinality allows that 1 PASODOBLE dependent variable be mapped as one element in a set of ISO keywords.
9	A.1	Identification and nature	Unit	For dimensionless variables, [value] = 'Dimensionless' unless the considered quantity is measured by a ratio (e.g. flat angle, relative humidity, mass or volume mixing ratio, etc.), involving or not involving a multiplication factor (examples: [value] = '%'; 'ppv'; 'ppmv'; 'ppbv'; 'µg/g'; etc.).						NC	y, but could be more easily done in a community profile	not common		
10	A.1	Identification and nature	Type of data sourcing (high level)	[value] = 'Observation' / 'Derived from indirect measurement' / 'Combination of measurements' / 'Combination of measurement and simulation' (data assimilation) / 'Model simulation'. NB Data sources are described in more detail under Section B - Product generation.						NC	y	not common		
11	A.1	Identification and nature	Immediacy	[value] = 'Data record' / 'Running NRT' / 'Forecast'.						NC	n	not common		
12	A.1	Identification and nature	Spatial representativeness	Point data / summed or averaged in space (e.g. along height, longitude; over area).						NC	y, but more atomic in ISO 19115:2014	not common		
13	A.1	Identification and nature	Temporal representative	Instantaneous / averaged over time.						NC	y, but more	not common		

			ness								atomic in ISO 19115:2014			
14	A.1	Identification and nature	Value range representativeness	Minimum, maximum, other…						NC	unknown	not common		
15	A.2	Domain of definition	Spatiotemporality	[value] = 'Spatiotemporal' if the product (dataset or model) is defined over a portion of a subspace of the 4-D geophysical space. [value] = 'Not spatiotemporal' otherwise, i.e. if the independent variables are not horizontal, vertical and/or time coordinates. If the product cannot be described in terms of variables, the answer would be "not applicable".						NC	y	not common		
16	A.2	Domain of definition	Geographic bounding box	Irrespective of the product spatiotemporality, the product may be defined or valid over a certain (possibly not connex) geographic area – if the product is a geo-referenced spatiotemporal dataset, the area is its geographic domain of definition. The geographic bounding box is the tightest longitude/latitude spherical rectangle circumscribing this area. It is given by its westbound and eastbound longitudes, and southbound and northbound latitudes (in decimal degrees with a precision of at least two decimals).	4.1		identificationInfo/SDN_DataIdentification/extent[1...n]/EX_Extent/geographicExtent/	1...n	1:1			EX_GeographicBoundingBox	geographic position of the resource	Also possibility to define several disjoint boxes
17	A.2	Domain of definition	Geographic resolution	If relevant, the geographic resolution of the product should be provided as a set of distances or map equivalent scales. If the product is a geo-referenced spatiotemporal dataset, this piece of metadata will possibly repeat information provided below for the horizontal independent variable(s).	6.2		identificationInfo/SDN_DataIdentification/spatialResolution[0...n]/MD_Resolution/	1	1:1			distance	ground sample distance	Any unit of a controlled vocabulary can be used

18	A.2	Domain of definition	Number of independent variables	If the number of independent variables (IV) is equal to n, the attributes Name, Unit, Boundaries, Resolution will be provided below for each of the n IV (in the example, three times).			SpatialRepresentationInfo[0...n]/MD_GridSpatialRepresentation/axisDimensionProperties				y	can be gathered from the number of axisDimensionProperties within spatialRepresentationInfo	name of the axis	The MD_DimensionNameTypeCode lists all the possible axis types: x, y, z, track, line, sample, time
19	A.2	Independent Variable	Name	/			SpatialRepresentationInfo[0...n]/MD_GridSpatialRepresentation/axisDimensionProperties[1...n]/MD_Dimension/					dimensionName		In CDI-ISO 19139, the axes can only provide partial information on the variables on which dependent variables are measured against, eg.axes can be of type x, y, z, track, line, sample, time. For primary observation data, this is not necessarily a problem.
20	A.2	Independent Variable	Unit	/			SpatialRepresentationInfo[0...n]/MD_GridSpatialRepresentation/axisDimensionProperties[1...n]/MD_Dimension/resolution/					Measure@uom	degree of detail in the grid dataset	Why relative mapping: The unit used for the axis resolution is not necessarily the one used for the independent variable itself.
21	A.2	Independent Variable	Boundaries	Limits of the definition interval (linear bounding box) or series of intervals. For example: S-N limits of a group of cities or areas, a country, a region (e.g. a coastal region); start and end dates of a record, running NRT or running forecast; duration of a theoretical scenario; start date of an ongoing record; ...			identificationInfo/SDN_DataIdentification/extent/					EX_Extent		This is in part mappable to CDI-ISO 19115 by the extent, ie. the vertical, horizontal and temporal extent. These are not per se the boundaries for a single variable as in PASODOBLE. E.g. the geographic bounding box is represented by the boundaries of the x and y axis.

22	A.2	Independent Variable	Resolution	E.g. value of linear / logarithmic discrete interval; or number of subdivisions of the interval defined above; or the axis scale definition (e.g. t, t+6h, t+12h, t+18h); etc.			SpatialRepresentationInfo[0...n]/MD_GridSpatialRepresentation/axisDimensionProperties[1...n]/MD_Dimension/resolution/	1	1:1			Measure	degree of detail in the grid dataset	The Measure resolution is represented by both unit and value.
23	A.3	Domain of values	Confidence level	Probability that a data value belongs to the confidence interval (e.g. 90%).						NC	n	not common		
24	A.3	Domain of values	Confidence interval	Estimated boundaries of an interval including most data values (the probability that a data value is included in the confidence interval is equal to the confidence level set up above). Expressed in the same unit as the data unit.						NC	n	not common		
25	A.4	Quality	Existence of quality assessment	If the product quality is assessed in any way. [value] = 'Validated' / 'Not validated'	6.1					NC	y	not common		Quality in CDI-ISO 19139 is only assessed in terms of conformance (ConformanceResult) to a standard specification (result can be pass or no pass). Note that the CDI profile uses a very limited subset of the data quality elements from ISO 19115. This limited set fits to the PASODOBLE concept put forward in A.7 References.
26	A.4	Quality	Validation method	Where applicable, description of how the product is validated						NC	y	not common		
27	A.4	Quality	Validation dataset(s)	Citation and description of dataset(s) used for validation						NC	y	not common		
28	A.4	Quality	Validation result	Summary of product performance against defined criteria	6.1					NC	y	not common		
29	A.4	Quality	Content analysis	Summary of content analysis						NC	y	not common		

30	A.4	Quality	Number of quality indicator(s) (QI)	If the number of quality indicators (QI) is equal to n, the attributes Name, Definition, Unit, Relevance, etc. will be provided below for each of the n QI (in the example, one time).						NC	y	not common		
31	A.4	Quality Indicator	QI name	/						NC		not common		
32	A.4	Quality Indicator	QI definition	Definition and / or calculation formula. For a model or service, definition of the QI measuring the product performance.						NC	y	not common		
33	A.4	Quality Indicator	QI unit	For dimensionless QI, [value] = 'Dimensionless' unless the QI is a ratio involving a multiplication factor.						NC	n	not common		
34	A.4	Quality Indicator	QI relevance	[value] = 'Individual' if the QI is defined for each data point (e.g. absolute or relative error, etc.). [value] = 'Global' if the QI is defined for the whole dataset (i.e., it assumes a unique value), irrespective of the determination method (i.e. whether determined from a sample, from the entire population, from error budget considerations, etc.). NB A global QI may be derived from individual QI (e.g. the mean or maximum uncertainty, etc.).						NC	n	not common		
35	A.4	Quality Indicator	QI confidence interval	Estimated boundaries of an interval including most QI values (the probability that a QI value is included in the confidence interval is equal to the confidence level set up below). Expressed in the same unit as the QI unit.						NC	n	not common		
36	A.4	Quality Indicator	QI confidence level	Probability that a QI value belongs to the confidence interval (e.g. 90% in the above example).						NC	n	not common		
37	A.4	Quality Indicator	QI value	/						NC	n	not common		
38	A.5	History	Date of creation	/	5 (5.4)					NC	y	not common		

39	A.5	History	Version number	/	6.1		identificationInfo/SDN_DataIdentification/citation/CI_Citation/	0...1	1:1			edition	version of the cited resource	
40	A.5	History	Version issue date	/	5 (5.3)		identificationInfo/SDN_DataIdentification/citation/CI_Citation/	0...1	1:1			editionDate	date of the edition	
41	A.5	History	Change between current and previous versions	/	6.1					NC	y	not common		
42	A.6	Presentation, coding and format	Display	Table of numbers / Picture / etc.						NC	y	not common, as CDI only deals with datasets		
43	A.6	Presentation, coding and format	Computer format(s)	ASCII / binary file; NetCDF / HDF / etc.; Text / PDF; JPEG / PNG / etc.			distributionInfo/MD_Distribution/distributionFormat[1...n]/MD_Format/	1	1:1			name	name of the resource format(s)	The data format of the data product is provided in the distributionFormat element (of multiplicity 1...n) of MD_Distribution, which is an instance of MD_Format. The version of the format is also included. Why relative mapping: the computer format registered in CDI-ISO 19115 is not necessarily the computer format(s) provided by the service, because the system can map the data from each NODC to an autogenerated odv file.
44	A.6	Presentation, coding and format	Computer language	Computer language used within the product (for a model, the language of the code)						NC	y	not common, as CDI only deals with datasets		
45	A.6	Presentation, coding and format	Language(s)	Natural language used within the product	1.7		identificationInfo/SDN_DataIdentification/	1	1:1			language	language(s) used within the resource	The languages used within the resource is part of the MD_DataIdentification Class
46	A.6	Presentation, coding and format	Unique resource identifier (URI)	/	1.5		identificationInfo/SDN_DataIdentification/CI_Citation/identifier/MD_I	1	1:1			code	alphanumeric value identifying an instance in the	

							dentifier/						namespace	
47	A.7	References	Standard(s) of reference	Standard(s) applicable to the product, the product specifications or the product validation	7.1		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/specification/CI_Citation/	1	1:1			title	name by which the cited resource is known	The data quality procedures at any data lifecycle level (eg. validation by curator/derivation by developer/provision by provider) in CDI-ISO 19139 are solely expressed in terms of adherence to a standard (which must be a published resource). A conformance consists of a specification, an explanation and a pass value. The lifecycle level contexts themselves have no dedicated field, although they could be put in DQ_ConformanceResult/explanation. The ISO concept 'DQ_DataQuality/scope' could express this concept, although in the CDI implementation, the possible scopes are limited to dataset or data series.
48	A.7	References	Degree of conformity to the standard	For each standard quoted above	7.2		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/	1	1:1			pass	Indication of the conformance result where 0=fail and 1=pass	Why relative mapping: boolean vs. categorical variable
49	A.7	References	Publication(s)	Reference(s) of publication(s) describing, using or referring to the product or its validation			identificationInfo/SDN_DataIdentification/additionalDocumentation/SDN_Citation/					title	collects bibliographic references to the dataset, such as articles and related publications	

50	B.1	Developer(s)	Product developer(s)	Organisation(s) involved in the development of the product	9.1 + 9.2		identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/organisationName/	0...1 (mandatory when no individualName is provided)	1:1			SDN_EDMOCode	This codelist is a ISO version of the SeaDataNet list EDMO defining “European Directory of Marine Organisations”.	This codelist lists all oceanographical institutes by their EDMO code. CI_ResponsibleParty with role 'originator'
51	B.1	Developer(s)	Contact person(s)	/			identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/	0...1 (mandatory when no organisationName is provided)	1:1			individualName	name of the responsible person-surname, given name, title separated by a delimiter	CI_ResponsibleParty with role 'originator'
52	B.1	Developer(s)	Contact e-mail address(es)	/	9.1		identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/contactInfo/CI_Contact/address[1...n]/CI_Address/	1...n	1:1			electronicMailAddress	address of the electronic mailbox of the responsible organisation or Individual	CI_ResponsibleParty with role 'originator'
53	B.1	Developer(s)	Contact telephone #	/			identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/contactInfo/CI_Contact/phone[0...n]/CI_Telephone/	0...n	1:1			voice		CI_ResponsibleParty with role 'originator'
54	B.1	Developer(s)	Programme or project name(s)	Project(s) under the umbrella of which the product was developed or is made available			identificationInfo/SDN_DataIdentification/descriptiveKeywords[1...n]/MD_Keywords/keywor	1...n	1:1			SDN_EDMERPCode		This codelist lists all projects by their EDMERP code. MD_Keywords with type 'project'

							d							
55	B.1	Developer(s)	Project principal investigator(s) (PI)	For each project			identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/	0...1 (mandatory when no organisationName is provided)	1:1			individualName		Why relative mapping: The project is just a keyword of type 'project', no other entities can be couples to this. CI_ResponsibleParty with role 'principalInvestigator'
56	B.1	Developer(s)	PI's affiliation(s)	For each PI	9.1 + 9.2		identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/organisationName/	0...1 (mandatory when no individualName is provided)	1:1			SDN_EDMOCode	This codelist is a ISO version of the SeaDataNet list EDMO defining "European Directory of Marine Organisations".	This codelist lists all oceanographical institutes by their EDMO code. CI_ResponsibleParty with role 'principalInvestigator'
57	B.1	Developer(s)	PI's e-mail address(es)	For each PI	9.1		identificationInfo/SDN_DataIdentification/citation/CI_Citation/citedResponsibleParty[0...n]/CI_ResponsibleParty/contactInfo/CI_Contact/address[1...n]/CI_Address/	1...n	1:1			electronicMailAddress		CI_ResponsibleParty with role 'principalInvestigator'
58	B.2	Derivation chain and intermediate validation	Derivation chain stages	Successive intermediate products and algorithms used to transform a stage to the next one. The information below will be provided for each stage (including final product)	6.1		gmd:dataQualityInfo/gmd:DQ_DataQuality/gmd:lineage/gmd:LI_Lineage/	1	1:1			statement		The lineage in CDI-ISO 19115 is a string expressing information about either events or source data used in constructing the data specified by the scope, or lack of knowledge about

															the lineage. There are no rules on how to delimit stages in the string.
60	B.2	Stage	Product name	/						NC	y	not common			
61	B.2	Stage	Product description	Description of resulting product (for a dataset, level, grid, etc.)						NC	y	not common			
62	B.2	Stage	Source(s)	Description of source product, including the following (for intermediate product, reference to previous stage). Nature, coverage, version, origin, reference citation. Model / instrument / platform name, location, operator. Measurement technique.						NC	y	not common			
64	B.2	Stage	Algorithm(s)	Algorithm / model used to derive resulting product from source						NC	y	not common			
65	B.2	Stage	Product validation	Validation means including reference datasets, methods, QI and results. (For final product, reference to information provided under A.4)						NC	y	not common			
66	B.3	References	Standard(s) of reference	Standard(s) applicable to aspects of the product development	7.1		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/specification/CI_Citation/	1	1:1			title	name by which the cited resource is known	The data quality procedures at any data lifecycle level (eg. validation by curator/derivation by developer/provision by provider) in CDI-ISO 19139 are solely expressed in terms of adherence to a standard (which must be a published resource). A conformance consists of a specification, an explanation and a pass value. The lifecycle level contexts themselves have no dedicated field, although they could be put in DQ_ConformanceResult/explanation. The ISO concept

														'DQ_DataQuality/scope' could express this concept, although in the CDI implementation, the possible scopes are limited to dataset or data series.
67	B.3	References	Degree of conformity to the standard	For each standard quoted above	7.2		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/	1	1:1			pass	Indication of the conformance result where 0=fail and 1=pass	Why relative mapping: boolean vs. categorical variable
68	B.3	References	Publication(s)	Reference(s) of publication(s) describing the product development			identificationInfo/SDN_DataIdentification/additionalDocumentation/SDN_Citation/					title	collects bibliographic references to the dataset, such as articles and related publications	Why relative mapping: there is a distinction between the dataset itself and the validation steps that operated on it. Most likely the SDN developers mean that the subject of this element is "the mature dataset", so the dataset plus the validation steps.
69	C.1	Provider	Service name	/						NC	y	not common		The service name is implicit in CDI-ISO 19115, ie. 'SeaDataNet CDI Access Service'
70	C.1	Provider	Service provider	/	9.1 + 9.2		distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/organisationName	0...1 (mandatory when no individual Name is provided)	1:1			SDN_EDMO Code	This codelist is a ISO version of the SeaDataNet list EDMO defining “European Directory of Marine Organisations”.	This codelist lists all oceanographical institutes by their EDMO code.
71	C.1	Provider	Provider acronym	/						NC	n	not common		

72	C.1	Provider	Provider country	/			distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/contactInfo/CI_Contact/address/CI_Address/country	1	1:1			SDN_CountryCode	country of the physical address	This codelist lists all countries.
73	C.1	Provider	Contact person	/			distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/	0...1 (mandatory when no organisationName is provided)	1:1			individualName	name of the responsible person- surname, given name, title separated by a delimiter	
74	C.1	Provider	Contact e-mail	/	9.1		distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/contactInfo/CI_Contact/address[1...n]/CI_Address	1...n	1:1			electronicMailAddress	address of the electronic mailbox of the responsible organisation or Individual	
75	C.1	Provider	Contact telephone #	/			distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/contactInfo/CI_Contact/phone[0...n]/CI_Telephone/	0...n	1:1			voice	telephone number by which individuals can speak to the responsible organisation or individual	
76	C.1	Provider	Service operation start date	/	5 (5.2)					NC	y	not common		
77	C.2	Applications	Service purpose	/						NC	y	not common		The service is implicit, ie. SeaDataNet CDI Access Service
78	C.2	Applications	Targeted application(s)	/						NC	n	not common		The application is implicit, ie. SeaDataNet CDI Access Service
79	C.2	Applications	Number of example(s)	If the number of applications is equal to n, the attributes Description, User						NC	n	not common		

				name, User acronym will be provided below for each of the n applications (in the example, one time)										
80	C.2	Example	Description	/						NC	n	not common		
81	C.2	Example	User name	/	9.1 + 9.2					NC	n	not common		
82	C.2	Example	User acronym	/						NC	n	not common		
83	C.2	Example	User e-mail address	/	9.1					NC	n	not common		
84	C.3	Maintenance and update	Update frequency	Frequency at which new data are incorporated to the product.			identificationInfo/SDN_DataIdentification/resourceMaintenance[0...n]/MD_MaintenanceInformation/maintenanceAndUpdateFrequency/	1	2:1			MD_MaintenanceFrequencyCode	Frequency with which changes and additions are made to the resource after the initial resource is completed.	This codelist lists update frequencies (continually, daily, notPlanned,...). Note that the CDI-ISO 19115 field comprises both maintenance updates and data updates.
85	C.3	Maintenance and update	Update delay	Mean time between data production and availability.						NC	y	not common		
86	C.3	Maintenance and update	Revision frequency	Frequency at which the data product is reassessed.			identificationInfo/SDN_DataIdentification/resourceMaintenance[0...n]/MD_MaintenanceInformation/maintenanceAndUpdateFrequency/	1	2:1			MD_MaintenanceFrequencyCode	Frequency with which changes and additions are made to the resource after the initial resource is completed.	This codelist lists update frequencies (continually, daily, notPlanned,...). Note that the ISO field comprises both maintenance updates and data updates (the latter both "small" and "large", eg. Revision, reassessment)
87	C.3	Maintenance and update	Backup frequency	/						NC	y	not common		
88	C.3	Maintenance and update	Archive maintenance	[value] = ‘Archive maintained’ / ‘No archive maintained’						NC	y	not common		
89	C.3	Maintenance and update	Archive coverage	Time period covered by the archived data.						NC	n	not common		
90	C.3	Maintenance and update	Archive availability	Is the maintained archive partially or totally available? [value] = ‘Archive available’ / ‘No archive available’						NC	n	not common		
91	C.3	Maintenance and update	Available archive time coverage	Time period covered by the available historic data.						NC	y	not common		

92	C.4	Delivery	Delivery mode(s)	[value] = 'Push mode' / 'Pull mode' / 'Consultation'			distributionInfo/MD_Distribution/transferOptions[1...n]/MD_DigitalTransferOptions/onLine[1...n]/CI_OnlineResource/function/	0...1	2:1			CI_OnLineFunctionCode	code for function performed by the online resource	This codelist lists the functions the service is capable of (download, information, offlineAccess,...) They can encompass both the way in which the service is delivered as their terms&conditions
93	C.4	Delivery	Delivery channel(s) or medium(s)	Internet download / FTP / E-mail / SMS / CD / etc.			distributionInfo/MD_Distribution/transferOptions[1...n]/MD_DigitalTransferOptions/onLine[1...n]/CI_OnlineResource/	0...1	1:1			protocol	connection protocol to be used e.g. http, ftp, file	Why relative mapping: the delivery possibilities are not protocols in the strict sense
94	C.4	Delivery	URL	/	1.4		distributionInfo/MD_Distribution/transferOptions[1...n]/MD_DigitalTransferOptions/onLine[1...n]/CI_OnlineResource/	1	1:1			linkage	location (address) for on-line access using a Uniform Resource Locator/Uniform Resource Identifier address or similar addressing scheme	
95	C.4	Delivery	Delivery frequency							NC	y	not common		
96	C.4	Delivery	Delivery delay	Mean time between product request and provision.						NC	y	not common		
97	C.5	Terms and conditions	Usage conditions or restrictions	E.g. non commercial, non military, acknowledgement of provider, offer of co-authorship, etc.	8.1				1:2			identificationInfo/SDN_DataIdentification/resourceConstraints[1...n]/MD_LegalConstraints/accessConstraints[0...n]/MD_RestrictionCode + /identificationInfo/SDN_DataIdentification/resourceConstraints[1...n]/MD_Le	access constraints applied to assure the protection of privacy or intellectual property, and any special restrictions or limitations on obtaining the resource or Metadata	Constraints not present in the MD_RestrictionCode can be represented in the otherConstraints

												galConstraints/otherConstraints[0...n]		
98	C.5	Terms and conditions	Access conditions	E.g. fee, password protection, online agreement with terms and conditions, signature of agreement, etc.	8.1 + 8.2		distributionInfo/MD_Distribution/transferOptions[1...n]/MD_DigitalTransferOptions/onLine[1...n]/CI_OnlineResource/function/	0...1	2:1			CI_OnLineFunctionCode	code for function performed by the online resource	This codelist lists the functions the service is capable of (download, information, offlineAccess,...) They encompass both the way in which the service is delivered as their terms&conditions
99	C.6	Support	Language(s) used to provide support	/						NC	n	not common	Although there are many roles possible in the ISO scheme, for this exercise the support is surmosed to be performed by the provider (e.g. distributor)	
100	C.6	Support	Additional service(s) provided	[value] = 'Instructions' (on format, software, ...) / 'Search engine' / 'File format reader' / 'File format converter' / 'Data extractor' (subsetting) / 'Visualisation' (plotting) / 'Data handling' / 'Alert' / 'Help desk' / 'Other service'						NC	y	not common		

101	C.6	Support	Alternative formats that can be automatically generated	If a file format converter is provided, which formats does it handle?						NC	y	not common		
102	C.6	Support	Data handling options provided	Types of data manipulations allowed online, if any. Examples: averaging, interpolation, comparison with other datasets, difference calculation, etc.						NC	y	not common		
103	C.6	Support	Help desk telephone	/			distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/contactInfo/CI_Contact/phone[0...n]/CI_Telephone/	0...n	1:1			voice	telephone number by which individuals can speak to the responsible organisation or individual	Why relative mapping: the support isn't necessarily performed by the distributor
104	C.6	Support	Help desk e-mail	/			distributionInfo/MD_Distribution/distributor/MD_Distributor/distributorContact/CI_ResponsibleParty/contactInfo/CI_Contact/address[1...n]/CI_Address	1...n	1:1			electronicMailAddress	address of the electronic mailbox of the responsible organisation or Individual	Why relative mapping: the support isn't necessarily performed by the distributor
105	C.6	Support	Other	Description of other service(s) provided						NC	y	not common		
106	C.7	References	Standard(s) of reference	Standard(s) applicable to the development of the services associated to the product						NC	n	not common		
107	C.7	References	Publication(s)	Reference(s) of publication(s) describing the services associated to the product						NC	n	not common		The CDI-ISO 19115 dataQualityInfo and additionalDocumentation

														elements only apply to the dataset and its validation.
108	D.	Metadata	Author(s)	Organisation(s) responsible for the creation and maintenance of the metadata	10.1		contact/CI_ResponsibleParty/organisationName/	0...1 (mandatory when no individual Name is provided)	1:1			SDN_EDMO Code	This codelist is a ISO version of the SeaDataNet list EDMO defining “European Directory of Marine Organisations”.	This codelist lists all oceanographical institutes by their EDMO code.
109	D.	Metadata	Contact e-mail address(es)	For queries about the metadata	10.1		contact/CI_ResponsibleParty/contactInfo/CI_Contact/address[1...n]/CI_Address/	1...n	1:1			electronicMailAddress	address of the electronic mailbox of the responsible organisation or Individual	
110	D.	Metadata	Date of creation	Date of creation of the metadata	10.2		root	1				dateStamp	date that the metadata was created	
111	D.	Metadata	Date of last update	Date of last update of the metadata (e.g. in the case of dynamical metadata)	10.2					NC	y	not common		
112	D.	Metadata	Language	Language in which the metadata are expressed	10.3		root	1	1:1			language		
113	D.	Metadata	Standard(s) of reference	Standard(s) applicable to the metadata	7,1		root	1	1:1+ (2)			MetadataStandardName + metadataStandardVersion	name of the metadata standard (including profile name) used + version of the metadata standard (version of the profile) used	Why relative mapping: only for the main metadata standard (note the multiplicity of 1). In CDI-ISO 19115 this text field is limited to 'ISO 19115/SeaDataNet profile'.
114	D.	Metadata	Standard(s) of reference	Standard(s) applicable to the metadata	7.1		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/specification/CI_Citation/	1	1:1+ (2)			title	name by which the cited resource is known	Why relative mapping: what additional model the metadata model adheres to and to what degree can be approximated in CDI-ISO 19115 by using by using the DQ_ConformanceResult class.

115	D.	Metadata	Degree of conformity to the standard	For each standard quoted above	7.2		dataQualityInfo/DQ_DataQuality/report[1...n]/DQ_DomainConsistency/result[1...2]/DQ_ConformanceResult/	1	1:1			pass	Indication of the conformance result where 0=fail and 1=pass	Why relative mapping: boolean vs. categorical variable / only for the 'external' metadata standard.