NZ Marine Geospatial Information Metadata Guideline

The metadata guideline provides recommendations on the content, structure and formats of metadata describing marine geospatial datasets in New Zealand to ensure consistency in how marine geospatial data is described, managed, published and therefore to facilitate its discovery.

**Structure, field names and contents**

The following table describes the field name and field content conventions a data provider shall use when creating, managing, and providing metadata for a dataset.

* For simplicity, a ‘flat’ structure is followed (as opposed to a hierarchical approach). This allows implementation as a simple tabular format.
* Some of the fields only require ‘Text’ entry and it is left to the data provider to determine the most useful content according to the guidelines in column ’Description / Guidelines for field contents’. No field size limit is prescribed.
* For some of the fields, content best practices are recommended. Using these content best practices will enable easy discovery / enterprise search across published data sources. Guidance for common formats (e.g. use of ISO8601 datetime string for Temporal Coverage) is provided in a separate table below.
* This guideline specifies minimum mandatory fields, and as a best practice a data provider shall aim to provide as many metadata fields as possible and useful (to comprehensively describe the dataset). A data provider can add more data fields for this purpose, and, preferably these additional fields shall be based on existing standards, such as Ecological Markup Language (EML), Dublin Core, ISO19115 or other relevant published standards.

**Relationship to relevant standards**

This guideline / best practice shall be consistent with existing metadata standards, and a range of relevant (international) standards that are already heavily used are referred to here, and equivalent fields / structures for these standards are listed.

* Ecological Markup Language (EML), as used in GBIF / OBIS for describing ecological datasets <https://eml.ecoinformatics.org>
* ISO19115-3 International Standard provided through the International Standards Organisation (ISO) for ’Geographic information — Metadata’ <https://www.iso.org/standard/32579.html>
* The Dublin Core™ Metadata Element Set is a vocabulary of fifteen properties for use in resource description maintained by the Dublin Core™ Metadata Initiative (DCMI) <https://www.dublincore.org/>

**Use of vocabularies**

Use of vocabularies is important to enable consistent search and discovery across published data sources. However, mandated vocabularies limit flexibility in describing specifics / details of data, and marine geospatial data is diverse in nature. Here the following approach is used to enable maximum flexibility.

* For relevant fields the use of vocabularies is recommended in column ‘Format’.
* Generally, the option is left open for NOT using a vocabulary but using ‘free text’.
* Vocabularies for use are recommended under column ‘Vocabulary options’. Users can choose one (or more for fields with multiple entries) of these vocabularies to use. References and URLs for these ‘Vocabulary options’ is provided in the “Vocabulary option” table below.
* Users can elect to use other vocabularies, and preferably these shall be published vocabularies.
* If a vocabulary is used, the reference to the published vocabularies shall be included, preferably through a persistent URL/URN.
* For some core fields vocabularies are mandated to ensure enterprise search / discovery is possible.

Metadata guideline

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Field** | **Mandatory****(Y/N)** | **Multiple entries****(Y/N)** | **Description/ Guidelines for field contents** | **Format** | **Format guidelines/ examples** | **Vocabulary options** | Dublin Coreequivalent | Ecological Markup Languageequivalent | ISO19115equivalent |
| **General Information** |
| Title | Y | N | Title for the dataset. Should be not more than several words, short sentence. | Text | - | - | //dc /terms /title | //eml /dataset /title | //mdb:MD\_Metadata /mdb:identificationInfo//mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation /cit:title |
| Description | Y | N | Description of the dataset shall cover summary information on the why, what, and how of the dataset. This should include following elements.Purpose of the data collection;Used methods and/or protocols; Broad spatial coverage; Time period of data collection;Types of data;Taxonomic coverage (if applicable);Other relevant descriptive information;Note: This replicates some field contents – but provides important summary information in one place. | Text | - | - | //dc /terms /description | //eml /dataset /abstract | //mdb:MD\_Metadata /mdb:identificationInfo//mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation /cit:abstract |
| Theme | Y | N | Theme that best describes the content of the data. Theme aims to help filtering / searching data of interest. | Vocabulary preferred | - | NZ MGI Themes | - | - | - |
| Type | Y | N | Type of data that best describes the content of the data. The type of data aims to narrow down the filtering / searching of data of interest. | Vocabulary preferred | - | NZ MGI Themes | - | - | - |
| Keywords | Y | Y | Keywords that best describe the content of the data.  | Vocabulary preferred with optional free text | - | NZ MGI ThemesAODN Parameter Category VocabularyNASA/Global Change Master Directory Earth Science Keywords | //dc /terms /subject | //eml /dataset /keywordset /keyword*AND* /keywordThesaurus | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:descriptiveKeywords /mri:MD\_Keywords /mri:keyword*WITH* /mri:thesaurusName*ALSO SEE* /mri:citation /mri:topic category |
| Licence | Y | N | A legal instrument governing the use constraints of the dataset. Typically, this should be accepted citation for a standard licence (such as CC) or be a link to a specific licence. | Text | Preferably, a standard licence framework shall be used, such as: Creative Commons, NZGOAL |  | //dc /terms /license | //eml /dataset /intellectualRights*WITH* /Licensed /licenseName*AND* /licenseURL | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:resourceConstraints /mco:MD\_LegalConstraints |
| Publication statement | N | N | Information about whether the data can be publicly released should it not be already opened. | Text | Y/N | - | - | - | - |
| Constraints | N | N | Any additional comments on potential restrictions / constraints on the use and usability of the data. | Text | - | - | //dc /terms /accessRights | //eml /dataset /additionalInfo | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:resourceConstraints /mco:useLimitation |
| Release date | N | N | Approximate date when the data can be released | ISO8601 format  | (See best practice for datetime string) | - | - | - | - |
| Citation | N | N | Citation the data provider wishes users to use. | Text | - | - | //dc /terms /bibliographicCitation | /eml /dataset /referencePublication*AND/OR* /usageCitation*AND/OR* /literatureCited | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation |
| Attribution | N | Y | Acknowledgement to the contributors (organisations) for the collection and/or hosting of the data, according to licensing terms.  | Text | - | - | Contributor? | ? | ? |
| Lineage | N | N | Contains a descriptive statement about the lineage/history of a dataset | Text | - | - | - | - | - |
| **Roles / Contacts (at the organisation level)** |
| Creator | N | Y | The creator are the organisation(s) responsible for the original creation of the dataset.  | Vocabulary preferred with optional Name String | (See best practice for Name String) | AODN Organisation Vocabulary | //dc /terms /creator | //eml /dataset /creator | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:pointOfContact*WITH* /cit:role /CI\_RoleCode @codeListValue=“author” |
| Owner | N | Y | Organisation(s) holding the rights for the dataset and (in cases where the dataset has been published) licensing it. | Vocabulary preferred with Name String | (See best practice for Name String) | AODN Organisation Vocabulary | //dc /terms /rightsHolder | //eml /dataset /associatedParty*WITH* /role=“owner” | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:pointOfContact*WITH* /cit:role /CI\_RoleCode @codeListValue=“owner” |
| Manager | N | Y | Organisation(s) responsible for the technical environment around maintaining the dataset, setting standards (and metadata) and ensuring safe storage of the dataset. | Vocabulary preferred with optional Name String | (See best practice for Name String) | AODN Organisation Vocabulary | NA | //eml /dataset /associatedParty*WITH* /role=“custodian” | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:pointOfContact*WITH* /cit:role /CI\_RoleCode@codeListValue=“custodian” |
| Provider | Y | N | Organisation responsible for providing the data (incl metadata).Note: This can potentially be different from ‘Manager’, e.g. where a different organisation publishes the dataset existence through a catalogue. | Vocabulary preferred with optional Name String | (See best practice for Name String) | AODN Organisation Vocabulary | //dc /terms /publisher | //eml /dataset /publisher*AND (for metadata)* */me*tadataProvider | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:pointOfContact*WITH* */ci*t:role /CI\_RoleCode@codeListValue=“publisher”*AND/OR* */*mdb:distributionInfo /mrd:MD\_Distribution /mrd:distributor /mrd:MD\_Distributor /mrd:distributorContact*AND (for metadata)*//mdb:MD\_Metadata /mdb:contact |
| Provider contact | Y | Y | Name, email and/or phone of the team/individual in organisation responsible for providing the data. | Text | - | - | - | - | - |
| **Spatial and Temporal Coverage** |
| Spatial Coverage | N | Y | Spatial rectangular bounding box specified by four coordinates. If a polygone (eg specific coordinates or bounding box) is not provided, a default bounding box will be assigned based on the geographic description provided to allow dataset display in a map viewer. | Decimal bounding box coordinates in simple features nomenclature | (See best practice for Simple Feature Well Known Text representation) | - | //dc /terms /spatial*AS* DCMI Box notation | //eml /dataset /coverage /geographicCoverage /boundingCoordinates | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:extent /gex:EX\_Extent /gex:geographicElement> /gex:EX\_GeographicBoundingBox |
| Coordinate Reference System | N | N | Spatial Reference System (SRS) used in describing Bounding Box | EPSG Code | (See best practice for spatial reference system (SRS) representation) | EPSG Code | NA | //emlr /dataset /spatialVector /spatialReference*OR* /spatialRaster /spatialReference | //mdb:MD\_Metadata /mdb:referenceSystemInfo /mrs:MD\_ReferenceSystem /mrs:referenceSystemIdentifier /mcc:MD\_Identifier /mcc:code |
| Geographic Description | Y | Y | Description of the study area, e.g. lake name, region, or similar | Vocabulary preferred with optional free text | - | NZ GazetterAODN Geographic Extents VocabularyGetty Thesaurus of Geographic Names | //dc /terms /spatial | //eml /dataset /coverage /geographicCoverage/geographicDescription | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:extent /gex:EX\_Extent /gex:description |
| Temporal Coverage | N | Y | Period of time over which data was captured. | ISO8601 format, in the case of open-ended date ranges for continuously updated datasets, where two dots (..) can be used. | (See best practice for datetime string) | - | //dc /terms /temporal*AS* DCMI Period | //eml /dataset /coverage /geographicCoverage /temporalCoverage | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:extent /gex:EX\_Extent /gex:temporalElement |
| **Generation / Methods** [Note that we follow the AODN model for vocabularies, which have ‘category’ and ‘detail’ fields to enable use of high-level controlled vocabs for easy discovery / search] |
| Generation method | N | N | This field allows for a text-based/human readable description of the method used for generating the dataset. The content of this element would be similar to a description of sampling procedures found in the methods section of a journal article. | Text | - | - | NA | //eml /dataset /methods | //mdb:MD\_Metadata /mdb:resourceLineage /mrl:LI\_Lineage /mrl:statement |
| Collection instrument type | N | Y | High level instrument category according to agreed vocabulary. This field enables to filter data collected in a similar technical way. | Vocabulary preferred with optional free text | - | ??? | NA | NA | //mdb:MD\_Metadata /mdb:acquisitionInformation /mac:MI\_AcquisitionInformation /mac:instrument /mac:MI\_Instrument /mac:type |
| Collection instrument | N | Y | Instrument make / model used in data collection.  | Vocabulary preferred with optional free text | - | AODN Instrument Vocabulary | NA | //eml /dataset /methods /methodStep /Instrumentation | //mdb:MD\_Metadata /mdb:acquisitionInformation /mac:MI\_AcquisitionInformation /mac:instrument /mac:MI\_Instrument /mac:identifier |
| Collection platform category | N | Y | High level platform type used for data collection, e.g. vessel, AUV, sensor platform to agreed vocabulary. This field enables to filter data collected in a similar technical way. | Vocabulary preferred with optional free text |  | AODN Platform Category Vocabulary | NA | NA | //mdb:MD\_Metadata /mdb:acquisitionInformation /mac:MI\_AcquisitionInformation /mac:platform /mac:MI\_Platform /mac:otherPropertyType |
| Collection Platform | N | Y | Specifics of platform used for data collection, e.g. make/model, vessel name, etc. | Vocabulary preferred with optional free text |  | AODN Platform Vocabulary | NA | NA | //mdb:MD\_Metadata /mdb:acquisitionInformation /mac:MI\_AcquisitionInformation /mac:platform /mac:MI\_Platform /mac:identifier |
| **Format / Storage** |
| Format type | N | Y | General description of the format of the dataset according to an agreed vocabulary: ascii, binary, … | Vocabulary preferred with optional free text | - | NASA/Global Change Master Directory Keywords Granule Data FormatsInternet Media Types [MIME] | //dc /terms /format | //eml /dataset /spatialVector /physical /dataFormat/externallyDefinedFormat*OR* /spatialRaster /physical /dataFormat/externallyDefinedFormat | //mdb:MD\_Metadata /mdb:distributionInfo /mrd:MD\_Distribution /mrd:MD\_Format /mrd:formatSpecificationCitation /cit:CI\_Citation /cit:title |
| Format  | N | Y | Specific description of data format for dataset; for example specific manufacture binary format reference. | Text | - | - | //dc /terms /format | //eml /dataset /spatialVector /physical /dataFormat*OR* /spatialRaster /physical /dataFormat | /mdb:MD\_Metadata /mdb:distributionInfo /mrd:MD\_Distribution /mrd:MD\_Format /mrd:formatSpecificationCitation /cit:CI\_Citation /cit:otherCitationDetails |
| **Content** |
| Parameter category | N | Y | This field describes in high level categories, the environmental parameters included with in dataset.High level parameter type/ category that applies to parameter to agreed vocabulary. This field enables to filter data of similar type. | Vocabulary preferred with optional free text | - | AODN Discovery Parameter Category Vocabulary | NA | NA | NA |
| Parameter | N | Y | This field describes the environmental parameters included with in dataset (e.g. ‘water temperature’) | Vocabulary preferred with optional free text | - | AODN Discovery Parameter Vocabulary | NA | NA | NA |
| **Identification / Versioning** |
| Identifier | N | N | Formal identifier of dataset, e.g. DOI, URI.Preferred is a persistent unique identifier for the dataset, provided through authoritative agency. | URI, URN, DOI, other industry standard identifier system | - | - | /dc /terms /identifier | //eml /datatset /@id | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation /cit:identifier |
| Version | N | N | Identifies the version of the dataset. When a dataset is published internally or externally it is assigned a new revision. The creation of this field is specific to organisation/ system/implementation. | Text | e.g. numeric 1.1, etc. | - | //dc /terms /hasVersion | NA (managed through managing full dataset version history) | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation /cit:edition |
| Date Created | N | N | Date/time that the dataset was created (This is not the surveyed date). For new datasets the current date will be used but may be edited. Publishing the dataset uses the internal date and may not be edited. (Note that some legacy datasets may not have this.) | ISO8601 format  | (See best practice for datetime string) | - | //dc /terms /created | NA (managed through managing full dataset version history) | //mdb:MD\_Metadata /mdb:identificationInfo /mri:MD\_DataIdentification /mri:citation /cit:CI\_Citation /cit:date*WITH* /cit:dateType /cit:CI\_DateTypeCode  @codeListValue=“creation” |
| Date Updated | N | N | Date/time that the dataset was updated and published internally. The current date will be used but may be edited. Publishing the dataset uses the internal date and may not be edited. If the first version this date/time will be the same created date/time. | ISO8601 format | (See best practice for datetime string) | - |  |  |  |
| Maintenance Frequency | N | N | Indicates how frequently an updated record/entry may be distributed to the publication platform. | ISO8601 format | (See best practice for datetime string) | - | ? | ? | ? |
| **URL** |
| Organisation URL | N | N | URL to the organisation website. | Text | - | - | - | - | - |
| Data URL | N | N | Link to the dataset itself, for which an HTTP GET request will initiate a download of the full dataset associated with the metadata file.  | Text | - | - | - | - | - |

Vocabulary options

|  |  |  |
| --- | --- | --- |
| **Vocabulary Name** | **URL** | **Notes** |
| AODN Organisation Vocabulary | <https://vocabs.ardc.edu.au/viewById/28>  | Extension may be required to apply to NZ |
| AODN Geographic Extents Vocabulary | <https://vocabs.ardc.edu.au/viewById/61>  | Extension may be required to apply to NZ |
| NZ Gazetter | <https://gazetteer.linz.govt.nz/>  | Extension may be required to apply to marine space |
| NASA/Global Change Master Directory Earth Science Keywords | <https://earthdata.nasa.gov/earth-observation-data/find-data/gcmd/gcmd-keywords>  |  |
| NASA/Global Change Master Directory Keywords Granule Data Formats | <https://earthdata.nasa.gov/earth-observation-data/find-data/gcmd/gcmd-keywords>  |  |
| DCMI Type Vocabulary | <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/#section-7>  | Probably does not apply here |
| AODN Parameter Category Vocabulary | <https://vocabs.ardc.edu.au/viewById/24>  |  |
| AODN Discovery Parameter Vocabulary | <https://vocabs.ardc.edu.au/viewById/22>  |  |
| AODN Platform Category Vocabulary | <https://vocabs.ardc.edu.au/viewById/26>  |  |
| AODN Platform Vocabulary | <https://vocabs.ardc.edu.au/viewById/25>  |  |
| AODN Instrument Vocabulary | <https://vocabs.ardc.edu.au/viewById/27>  |  |
| Getty Thesaurus of Geographic Names | <http://www.getty.edu/research/tools/vocabularies/tgn/index.html>  | Probably does not apply to NZ |
| Internet Media Types [MIME] | <https://www.iana.org/assignments/media-types/media-types.xhtml>  |  |
| NZ MGI Themes | <https://www.linz.govt.nz/file/23674/download?token=FmYE__8a>  |  |

Best practice/conventions for common field content

|  |  |
| --- | --- |
| **Best Practice for** | **Names String** |
| Examples | Best Practice examples: ‘Joe Bloggs’; ‘University of Waikato/Joe Bloggs’Not preferred: ‘JB’; ‘J. Bloggs’ |
| Description | This should describe as best as possible the relevant individual and /or institution responsible for the particular workflow in question.The full name of the individual should be given. If the name of individual is not available, the name of the institution only can be given.If the known individual is related to a known external organisation this can be indicated by ‘organisation name’/’individual name’. |

|  |  |
| --- | --- |
| **Best Practice for** | **Datetime String** |
| Examples | 2014-09-16T10:50+122014-09-162014-0920142007-03-01T13:00:00Z/2008-05-11T15:30:00Z |
| Description | ISO8601 shall be used to construct any datetime string.**References**<http://en.wikipedia.org/wiki/ISO_8601><http://www.iso.org/iso/home/standards/iso8601.htm>  |

|  |  |
| --- | --- |
| **Best Practice for** | **Simple Feature Well Known Text (WKT) representation** |
| Examples | POINT ( 48.4 175.6)LINESTRING(176.953 -41.768, 176.87 -41.624) |
| Description | Simple feature well known text representation shall be used, see the following references.**References**Open Geospatial Consortium (OGC): OpenGIS Implementation Specification for Geographic Information – Simple feature access – Part 1: Common Architecture. <http://www.opengeospatial.org/standards/sfa> ZM values and SRID for Simple Features. (PostGIS Extended Well-Known Text Format).<https://github.com/gravitystorm/postgis/blob/master/doc/ZMSgeoms.txt> ISO/IEC 13249-3:2011. Information technology -- Database languages -- SQL multimedia and application packages -- Part 3: Spatial. <http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=53698> International Association of Oil & Gas Producers (OGP) EPSG Geodetic Parameter Dataset. <http://www.epsg-registry.org/>  |

|  |  |
| --- | --- |
| **Best Practice for** | **Spatial reference system (SRS) representation** |
| Examples | EPSG:4326 |
| Description | Any spatial reference system (SRS) description shall be done in the form EPSG:xx, where xx is the EPSG code for the reference system according to the EPSG Geodetic Parameter Dataset.**References**International Association of Oil & Gas Producers (OGP) EPSG Geodetic Parameter Dataset. <http://www.epsg-registry.org/>  |