Instructions for the preparation of data packages for the Index of Marine Surveys for Assessments (IMSA)

Purpose of these instructions

To assist proponents in preparing data packages for the Index of Marine Surveys for Assessments (IMSA). The Environmental Protection Authority (EPA) and Department of Water and Environmental Regulation (DWER) require IMSA data packages to support assessment processes under the *Environmental Protection Act 1986* (EP Act).

Whenever a marine survey report is provided as part of the environmental impact assessment (EIA) process, the report and associated raw data must be provided electronically as part of an IMSA data package.

The EPA has issued a series of three technical guidance documents to ensure that adequate marine information is obtained and used in the EIA process. To determine what marine surveys need to be undertaken refer to the relevant EPA technical guidance:

- Technical Guidance - Protection of Benthic Communities and Habitats
- Technical Guidance - Protecting the Quality of Western Australia's Marine Environment
- Technical Guidance - Environmental Impact Assessment of Marine Dredging Proposals.

Each technical guidance document contains a section that describes the kind of information that is required to be submitted as part of the EIA process, as well as a reference to this document - *Instructions for the preparation of data packages for the Index of Marine Surveys for Assessments*.

These instructions outline the data requirements and submission process for IMSA data packages, and the options that Intellectual Property (IP) owners (that is owners of any intellectual property rights in the material) have regarding the public availability of IMSA data.

Note that these instructions and the accompanying electronic templates and forms will be updated as the IMSA data standards are refined – particularly during the first year of IMSA's operation. Users should consult the EPA website to ensure they have the most recent versions of these documents prior to using them.

IMSA data packages

Marine data packages are to be submitted to IMSA accompanied by the relevant marine survey report and bundled as part of an IMSA data package. IMSA data packages are only required for marine surveys that have not been previously captured in IMSA. This includes all new generated survey reports and data as well as previous survey reports and data (if available) included in desktop studies.

A single environmental review document (ERD) or other documentation provided to the EPA may be accompanied by one or more IMSA data package. Each IMSA data package should be associated with a single marine survey report (often included as an appendix to the ERD). Where a single marine survey report documents multiple field surveys, the data from the different surveys should be contained in multiple files in the single IMSA data package for that report.

An IMSA data package is a single file in zip format (excluding large processed / source data), containing:

- one **metadata and licensing statement** in pdf format (see Form: IMSA Data Package)
- one **survey report** in pdf format
- one **plain-text survey report** in txt format
- a set of electronic data files, comprising items listed in the IMSA data standard (section 1.1).

Each file within the IMSA data package should comply with the IMSA data standards provided in these instructions. In naming the files within the IMSA data package, the terms presented above in bold should be incorporated into the file names.

IMSA processes

Proponents are to submit IMSA data packages using existing EIA processes and systems. There is no requirement for proponents to directly upload information to IMSA. IMSA data packages should be treated as electronic appendices attached to referrals, environmental review documents and other submissions made to the EPA. Any changes or corrections made to the marine survey report or underlying data during assessment or a compliance process should be reflected in an updated IMSA data package submitted to the EPA or DWER.

Once an assessment process has concluded, regardless of any appeals processes that may subsequently occur, the metadata component of each IMSA data package, that is the metadata and licensing statement (see box one), will be made publicly available on the IMSA website. This will occur for all IMSA data packages, allowing all marine surveys conducted in Western Australia to be indexed and discoverable in IMSA, even if the datasets themselves are not made available.

Box one: Data versus metadata

Data are pieces of information, such as measurements, observations and counts. Metadata summarise information about data and provide context by describing aspects of how data was obtained. In the context of an IMSA data package, the data are the pieces of marine information collected in the field, such as records of benthic communities observed at specific times and places, or a resulting derived dataset such as a habitat map. There may be large amounts and different types of data in an IMSA data package, depending on the size and complexity of the survey. The metadata are basic details about the survey itself, such as when, where and by whom it was conducted.

IP owners have control over whether the data component of an IMSA data package, that is the survey report, plain-text survey report and electronic data files, will be made publicly available on the IMSA website. This control is exercised with the use of licences that are granted by IP owners as part of the metadata and licensing statement – the licences can allow the data component to be withheld, be published or be both published and made re-usable by third parties.

The explanatory notes accompanying the metadata and licensing statement explain these options in full. It is essential that proponents and environmental practitioners understand who the IP owner is before the metadata and licensing statement is completed (see box two).

Box two: Intellectual property (IP) rights

Proponents who intend to license their data for publication or re-use via the IMSA website must ensure that they own the IP rights to that material or that they otherwise have the right and authority to grant a licence. That is, they themselves must obtain all necessary licences and consents from any relevant IP owner. A clear understanding of IP rights is essential when proponents engage third parties to conduct marine surveys on their behalf. Proponents and environmental practitioners should discuss their expectations around data sharing as early as possible. Proponents should ensure that their requirements for intellectual property ownership and sharing are clearly defined during the procurement process or by formal agreement.

IMSA data standards

Collectively, the remainder of these instructions, the electronic IMSA templates and the metadata and licensing statement (with its accompanying notes and definitions) make up the IMSA data standards. The electronic IMSA templates and the metadata and licensing statement are available for download here.
The IMSA data standards define a minimum required set of attributes for each electronic data file. Electronic data files that contain additional attributes will be accepted.

IMSA data packages should be provided as a single zip file containing all required files (exceptions for large processed / source data). IMSA data packages that are incomplete, or do not otherwise comply with these instructions or the IMSA data standards, will be returned for correction.

1. IMSA Data Standard

Proponents will be required to submit the marine survey data and metadata collected in support of their environmental review or compliance documentation to the EPA/DWER to be made available to the public when the assessment outcome is published.

This section of the document outlines the specific requirements for the data submission and informs the corresponding policy changes.

Data/information is organised into five hierarchical levels:

1. report (for example, technical reports as appendix to environmental review document)
2. metadata/spatial index (for example, contains bounding box/polygon(s) of data location)
3. report map layers/interpreted data (for example, GIS layer(s), annotation spreadsheet)
4. standard processed products/standard grid data (for example, multibeam or LiDAR products)
5. source (for example raw video/photos).

IMSA initially will focus on four data categories; modelling inputs, marine fauna, benthic communities and habitat, and marine environment quality, as shown in Table 1.

Table 1: IMSA data categories

Data Category	Category Description	Relevant Data
Modelling input	This category contains the physical oceanographic data routinely collected for input into models.	Source >> hydrodynamics • currents (direction, speed, depth) • waves (period, plus) • oscillations • CTD (conductivity, temperature, depth) • date • source terms • validation data (confidence interval).
Marine fauna	This category contains any data collected about marine fauna.	Map layer >> marine fauna mapping • area • classification • seasonality • date • condition • critical environmental windows • time series • habitat.

Data Category	Category Description	Relevant Data
Benthic communities and habitats	This category contains any data collected about the seafloor habitat and associated communities.	Map layer >> habitat mapping • area • classification • seasonality • date • condition • critical environmental windows • time series. Map layer >> regional distribution • area • classification. Map layer >> local assessment unit • extent • area • rationale.
Marine environmental quality	This category contains any data collected about water, sediment and biota quality.	Map layer >> environmental quality plan • values • area • objectives • ecological protection. Standard product >> water quality • location • parameters • values • frequency • duration. Standard product >> sediment quality • location • parameters • values • frequency • duration. Standard product >> biota quality • location • parameters • values • frequency • duration.

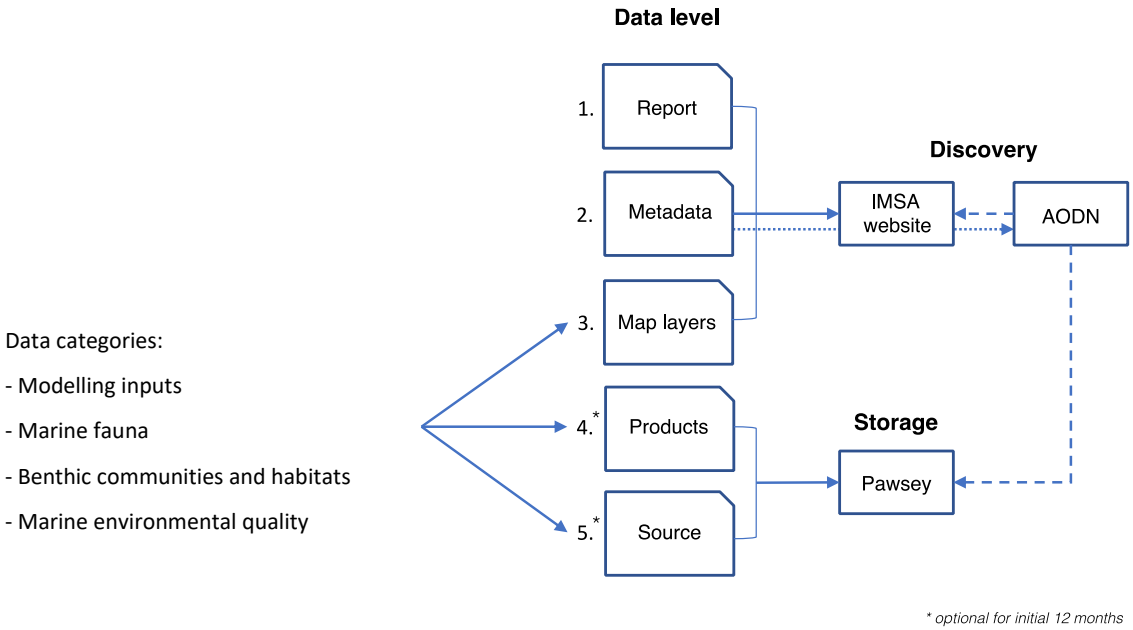

Figure 1: Model of proposed IMSA data workflow

Staged Implementation

In order to minimise the change effort required and to allow proponents to adjust to the new processes, the standard will be implemented in a staged process (Table 2) with progressively detailed data to be provided.

Table 2: Summary staged implementation

Deliverables	30 March 2020	30 March 2021

• Report metadata
  • *.pdf file of report
  • *.txt file of report (or *.xml)
  • quality control statement.
	Must comply with the Report Metadata Standard.	Must comply with the Report Metadata Standard.

• Survey dataset	Must be submitted in electronic format that conforms to level 3 and *optionally** comply with level 4 and 5. **optional for initial 12 months*	Provide all available data from level 3 to 5.

Proponents are encouraged to submit data with associated metadata as this will allow the data to be registered and indexed for future uses and discoverability. The standards referenced in Table 2 above are defined further in the following sections.

1.1 IMSA data capture standard

A minimum set of data standards have been drafted in relation to the capture and subsequent submission of marine survey data. Collectively, these standards are referred to as the IMSA survey data capture standards, and will apply to data submitted in support of environmental assessments under Part IV and Part V of the *Environmental Protection Act 1986* (EP Act).

The standard is comprised of the following core components:

- report metadata
- child metadata
- survey dataset.

These components are described in the following sections.

1.1.1 Report metadata

Proponents will be required to submit a minimum level of documentation relating to each marine survey report that is used in support of their environmental review and compliance documentation. This documentation is referred to as the report metadata.

The report metadata is designed to be sufficiently comprehensive so as to provide a useful and searchable record of the marine survey, while at the same time be simplistic enough to ensure the compliance to the standard is not onerous for proponents. The establishment of the standard was done with consideration of the following nationally and internationally recognised data and metadata frameworks:

- International Standards Organisation Geospatial Metadata Standards (ISO 19115) – marine community profile
- Industry-Government Environmental Meta-Database (I-GEM) metadata standards
- requirements of BioCollect, a component of the Atlas of Living Australia.

One of the primary drivers behind each of the above frameworks is the establishment of the 'who, what, where, when and how' aspects of collections of data. Commonality between these frameworks was considered in the context of marine survey data, and through this process the minimum metadata elements were derived to form the standard.

It is envisaged that the report metadata will accompany a submission to the regulator, via the completion of a simple pro-forma which encompasses the key mandatory metadata elements. The full standard can be found in Appendix A.

1.1.2 Child metadata

A child metadata standard has been established in recognition of the fact that often a single marine survey report can be comprised of information gathered from several surveys. These surveys may have materially different survey areas and related data, and as such the metadata related to each survey is required to be captured under this standard.

The specific metadata elements required under this standard are closely aligned to the report metadata standard discussed in the previous section.

1.1.3 Survey dataset

In addition to the metadata describing the report and associated surveys, there is also a requirement for proponents to submit data generated by the surveys, and referenced in the reports. It is through the submission of these datasets, and their subsequent publication to the marine community of Western Australia, that a valuable data asset for the marine environment will be developed over time.

The survey dataset has been developed to govern the format and structure of survey data. The standard describes the minimum fields which must be included and populated, with the expectation that proponents will include as many additional data fields as required to ensure all data is comprehensive and relevant to the submission.

See Appendix B for details.

2. Appendix A – metadata standard

Definition: The below standard defines the minimum information which must accompany any referral or application that is lodged with the EPA/DWER that includes a marine survey report. The metadata statement must contain both report and child metadata (if appropriate). The report metadata (or parent metadata record) contains details of the overall report and location while the child metadata record sits underneath this report metadata and contains details of individual surveys (with the same metadata fields as a report metadata record).

To aid in compliance, the IMSA metadata standard consists of the Index of Biodiversity Surveys for Assessments (IBSA) metadata fields plus some extra fields (which can be auto-populated) to ensure compliance with the Australia Ocean Data Network (AODN) metadata standard.

Timeframe: Level 1 – 3 data standards (see Figure 1) mandatory for all lodgements that occur on or after 30 March 2020.

Format: The submission of this metadata is envisaged to occur via a pro-forma document, to be populated by proponents. Fields, descriptions and examples can be found below.

Required fields:

Table 3: Report metadata (IBSA and AODN fields in blue, which some can be autogenerated). Fields in green to be added as required post upload to IMSA website

Field	Description	Example	Requirement
Title *(report name)*	The name of the survey report	Seagrass survey of Secret Sound	Mandatory
Abstract *(survey type(s))*	A description of the survey in the report	A benthic habitat survey of Secret Sound was conducted to measure percentage seagrass cover	Mandatory
Responsible party / role *(survey consultant)*	The name of the person or group who undertook the survey(s), typically the primary author(s) of the report	Fred Bloggs (fred.bloggs@abcconsulting.com.au)	Mandatory
Temporal extent *(start date)*	The date on which the earliest survey referenced in this report commenced	15/08/2018	Mandatory
Temporal extent *(end date)*	The date on which the last survey referenced in this report concluded (if known)	19/09/2018	Optional
Spatial extent *(encompassing survey boundary)*	An attached image, or reference to a figure (map) contained in the survey report, which depicts the encompassing boundary of all surveys included in this report	Coordinates	Mandatory
Supplementary information *(Citation)*	A citation of the accompanying survey report, to be referenced at the time of data publication	Seagrass Study of the East Kimberley, ABC Consulting (2017)	Mandatory
Citation date	Identification of when a given event occurred - CI_DateTypeCode <codeList> 'creation' 'publication' 'revision'	Publication, 2019-08-30	Mandatory

Field	Description	Example	Requirement
Revision date	Date identifies when the resource was examined or re-examined and improved or amended	2019-08-30	Conditional
File identifier	Unique identifier for the metadata file	1efcf057-f0a5-46fd-bb1e-9d87c556c429	Mandatory
Topic category	High-level geographic data thematic classification to assist in the grouping and search of available geographic datasets (ISO 19115 Topic Category)	oceans, biota, environment	Mandatory
Standard name	Name of the metadata standard	Australian Marine Community Profile of ISO 19115:2005/19139	Mandatory
Standard version	Version (profile) of the metadata standard used	2.0	Mandatory
Language	Language used for documenting metadata	English	Mandatory
Date stamp	Date that the metadata was created (YYYY-MM-DD)	2019-08-28	Mandatory
Resources constraints (creative commons constraint)	License for data	Creative commons 4.0 Attribute (CC BY 4.0) license.	Conditional
Online resource	Online source from which dataset can be obtained	https://data.pawsey.org.au/download/IMSA/CompanyA/XY.zip	Conditional

3. Appendix B - survey dataset standards

Definition: The below standard defines the minimum standard for marine survey data that is submitted to a regulator in support of a referral or application (Part IV or V of the EP Act). The standard defines the minimum structure of each dataset, with the expectation that the proponents will add as many additional data fields to ensure all data is comprehensive and relevant to their submission.

The standard includes those datasets which are expected to be produced as a result of marine surveys, and proponents are to select which datasets are most relevant to the survey being undertaken.
All data should be quality controlled for completeness, accuracy, and any errors before delivery.

Timeframe: Level 4 and 5 data standards optional for first 12 months, mandatory from 30 March 2021.

File formats:

1. Report

The report shall be provided as a Portable Document Format (.pdf) and Text Format (.txt). Having a .txt file aids in future reuse and data mining.

2. Metadata

The metadata shall be provided in both pdf and txt format, with the option of providing it as xml if appropriate. To enable discovery of IMSA data by a wide audience, metadata will be made available via the BioCollect IMSA website and the AODN metadata catalogue (https://catalogue.aodn.org.au/geonetwork/).

3. Map layers / interpreted data

Esri File Geodatabase (gdb) is the preferred GIS data format. However, GIS data can also be submitted in one of the following industry-standard GIS data formats if necessary:

- Esri Shapefile (shp et al)
- MapInfo TAB File (tab).

Within the geodatabase, metadata stored as xml needs to capture a 'data dictionary' detailing the information about the dataset such as the meanings of codes and full description of attributes.

The survey details spatial dataset should use the Geocentric Datum of Australia 2020 (GDA2020) datum and be projected using the appropriate Map Grid of Australia zone, unless this is inappropriate due to the dataset's extent being larger than a single zone. The coordinate system will be defined natively with the submitted spatial data files, such as via a prj file in the case of submission of an Esri Shapefile.

All spreadsheet data should be submitted as a CSV (Comma Separated Values) file as a minimum with well-described header row (first line). Data should be free of formatting and special characters where possible. Definitions for each column heading should be described within a separate 'readme'.txt document.

GIS / spreadsheet data attributes:
Recommended attributes for IMSA GIS/spreadsheet data

Table 4A: Attributes recommended for a marine fauna spatial dataset

Field	*GDB Field*	*Shp Field*	Type	Description	Example
Taxon	*Taxon*	*Taxon*	Text(100)	A description of the marine fauna identified with inclusion of scientific name preferable	Humpback whale (Megaptera novaeangliae)
Start date	*StartDate*	*StartDate*	Date/time	The date the sampling commenced at this site. Time as well if appropriate. Use ISO 8601 standard.	2017-08-16T13:12:11

End date	*EndDate*	*EndDate*	Date/time	The date the sampling concluded at this site. Time as well if appropriate. Use ISO 8601 standard.	2017-08-18T07:11:23
Habitat	*Habitat*	*Habitat*	Text(100)	A description of the habitat animal was observed in with reference to classification system used e.g. CATAMI	Substrate: Unconsolidated (soft)
Comment	*Comment*	*Comment*	Text(255)	Any additional information	

Table 4B: Attributes recommended for a benthic communities and habitat spatial dataset

Field	*GDB Field*	*Shp Field*	Type	Description	Example
Habitat type	HabitatType	HabitatTyp	Text(100)	A description of the benthic community and habitat with reference to classification system used e.g. CATAMI	Cnidaria: Corals: Stony corals: Branching
Start date	*StartDate*	*StartDate*	Date/time	The date the sampling commenced at this site. Time as well if appropriate. Use ISO 8601 standard.	2019-07-16T07:11:23
End date	*EndDate*	*EndDate*	Date/time	The date the sampling concluded at this site. Time as well if appropriate. Use ISO 8601 standard.	2019-07-16T07:11:23
Condition	*Condition*	*Condition*	Text(100)	Condition of the biota	Bleached / Dead
Comment	*Comment*	*Comment*	Text(255)	Any additional information	

Table 4C: "Sample" headers and descriptions for a marine environmental quality spreadsheet

Field	*GDB Field*	*Shp Field*	Type	Description	Example
Site ref	*SiteRef*	*SiteRef*	Text (50)	The primary reference code or number used to identify the site. The site reference derives from a particular numbering system set by the project that employs a meaningful structure.	RTU0198
Collected date time	*CollectedDate*	*CollectDat*	Date/time	The date and time of the collected sample, down to the second.	16/06/2001
Sample ID	*SampleID*	*SampleID*	Integer	A unique ID for the sample	12235
Lab sample no.	*LabSampleNo*	*LabSmpleNo*	Integer	A tracking number applied to the sample by the analytical laboratory where applicable.	13245
Depth measurement point	*DepthMeasurementPt*	*DpthMeasPt*	Text (50)	The measurement or reference point from which the depth of the sample was measured	Surface water level (SWL)

Field	*GDB Field*	*Shp Field*	Type	Description	Example
Sample depth	*SampleDepth*	*SampDepth*	Number	A text field describing the depth or depth range at which the sample was collected in metres. Depth may be a single value (for example 2.5) for grabs and in situ samples, or a range (for example 0–3.5) for samples taken over depth.	2.5
Collection method	*CollectionMethod*	*CollMethod*	Text (100)	The means by which the sampler was able to capture or collect the matrix in order to measure it. It is not the instrument or device used to collect the sample, but rather the methodology employed.	Sediment Grab
Collection device	*CollectionDevice*	*CollDevice*	Text (50)	The device that was used to capture or collect the matrix in order to measure it. It is not the instrument used to take the measurement.	Petite Ponar (9L)
Collection frequency	*CollectionFrequency*	*CollFreq*	Text (50)	The general frequency of the sampling	Once-off
Sample matrix	*SampleMatrix*	*SampMatrix*	Text (50)	The physical medium that was sampled or measured	Water
Sample comment	*SampleComment*	*SampCommnt*	Text (255)	Comments made about the sample by the sampler or data capture staff.	

Table 4D: "Reading" headers and descriptions for a marine environmental quality spreadsheet.

Field	*GDB Field*	*Shp Field*	Type	Description	Example
Sample ID	*SampleID*	*SampleID*	Integer	A unique ID for the sample (see Table 4C)	12235
Variable	*Variable*	*Variable*	Text(50)	The variable (the basic element or property measured). *Note in DWER Water Information Reporting the Variable and Units combine to make the field "Determinand"	N (tot) {TN, pTN}
Units	*Units*	*Units*	Text(50)	Standard reporting units for the sample matrix	mg/L
Reading value	*ReadingValue*	*ReadingVal*	Various	The value being reported can be a	5 - 10

				number, range, date, text	
Variable name	*VariableName*	*VarName*	Text(50)	The name or identifier for a constituent or property that is measured or estimated in a sample or field measurement. See "variable and analysis methods" for details	N (tot inorg)
Variable type	*VariableType*	*VarType*	Text(50)	A means of classifying variables into groups having similar chemical structure, end use, biological order, physical characteristics and/or measurement technique. See "Variable and Analysis Methods" for details	Isotope ratios
Variable qualifier name	*VariableNameQualifier*	*VarNamQual*	Text(50)	A property that moderates or qualifies a variable, such as specimen length ranges	
Analysis method code	*AnalysisMethodCode*	*AMCode*	Integer	The code assigned by a laboratory or the department to uniquely identify the analysis method used to obtain the reading. For field readings this is generally "Direct reading". See "variable and analysis methods" for details	3000
Measurement instrument	*MeasurementInstrument*	*Instrument*	Text(100)	The field or laboratory instrument used to obtain the measurement	
Standard units	*StandardUnits*	*StndrdUnts*	Text(50)	The standard units for the variable and sample matrix being measured	mg/L
Limit of reporting (LOR)	*LimitOfReporting*	*LOR*	Various	Record the minimum concentration of a substance in a sample that can be reliably detected by a laboratory.	0.01 mg/L
Quality code	*QualityCode*	*QualCode*	Integer	The quality code for the reading (see Table 10 of WIR Explanatory Notes)	5

For further detail and descriptions refer to the Department of Water and Environmental Regulation's Water information reporting (WIR) system Explanatory notes document, specifically section 6.1 - data reports. Also see the variable and analysis method lists. The aim is to align IMSA marine environmental quality data to existing water information reporting as far as practicable.

4. Standard processed products / standard grid data
See recommendations below.

5. Source data
Rather than mandatory requirements we have a list recommended guidelines that aid in interoperability and reusability for IMSA data categories.

Marine fauna

- The Department of the Environment and Energy Guidelines for biological survey and mapped data
- EPA Marine Fauna Environmental Factor Guideline.

Benthic habitat

- The National Environmental Science Programme (NESP) Biodiversity Hub has created standard operating procedures in the collection and analysis of monitoring data. This will allow synthesis between institutions, regions and times for data collected into the future.
- CATAMI Classification Scheme – part of the review of existing international and national classification schemes from the seamap classification page.

Oceanographic / modelling inputs

- AusSeabed's mission is to improve the awareness, coverage, quality, discoverability and accessibility of seabed mapping data through coordination and collaboration in the Australian region with a focus on bathymetry. There is an AusSeabed data hub project which is developing best-known practices and standards. Outcomes of this development should be followed for relevant data types where appropriate.
- IMO / AODN NetCDF standards (TBD).

Marine Environment Quality

- EPA Marine Environmental Quality – Environmental Factor Guideline
- DWER Water Information Reporting
  - http://wir.water.wa.gov.au/Pages/HelpandReferences.aspx
  - https://kumina.water.wa.gov.au/waterinformation/WIR/Reports/Publish/Explanatory%20Notes.pdf
- National Industry Guideline for water quality metadata (also relevant for hierarchical level 3).