1. Zip file information

This zip file contains a MEDIN data guideline in the xlsx spreadsheet format. Within the excel workbook there are 'Guidance' spreadsheets and 'Form' spreadsheets. The guidance spreadsheets provide full descriptions, controlled vocabulary or format recommendations, and examples for each of the metadata fields. The form sheets can be used to fill in the information.

2. Guidance on data guidelines

2.1 Background to Data Guidelines

The Marine Environmental Data and Information Network (MEDIN) are working towards creating a framework of consistent guidelines covering the major types of data collection undertaken in the marine environment around the UK. The principal benefits of this suite of guidelines are:

- Allowing contracting organisation(s) to easily specify a format that data should be returned in that can be readily used and includes all relevant attributes
- Provides a consistent format for contractors to work to (rather than a different format for each contract)
- Data can be readily exported to Data Archiving Centres and other users
- Instils good practice amongst users

Each guideline defines the data and information that must be stored with a particular data type to ensure it can be readily used and reused. As this type of information is specific for different data types, guidelines are developed for each type. This document describes one such format. Other guidelines can be accessed through the MEDIN website.

2.2 Using this Data Guideline

The data guideline is split into three sections:

A   General Metadata

B   Detailed Metadata

C   Data

These contain fields that need to be populated to allow the supplied data to be re-used effectively. Each field is either; mandatory, conditional or optional as indicated by M, C, or O respectively. Conditional means that the field must be completed if a value is known.

In the event that historical data which do not have all the necessary mandatory fields are being configured into this guideline, then it is permissible to use the following entry terms:

Term	Description
unknown	The correct value is not known to and not computable by the creator of this information. However a correct value probably exists.
inapplicable	There is no appropriate value. To be used in cases where metadata elements cannot be set null due to schema constraints.

In some cases it may be necessary to extend this guideline for a specific purpose such as a specific exchange of data between applications or to fulfil the needs of a specific project. This is permissible, however MEDIN advise that the broad structure and format is maintained and that where possible controlled vocabularies are used. As any extension to the structure and format may be useful for other organisations please inform MEDIN of further agreements.

The guidelines structure the three sections into; a summary of the information (Section 1) and; tables in Appendices A, B and C, which give detailed examples of how the guideline should be used. The tables contain a description, a controlled vocabulary and/or format requirement and examples of how the fields should be populated.

The different sections cover the following:

A   General Metadata

The General Metadata tables are common to all Data Guidelines and so only need to be completed once for a survey even if a number of different techniques and data guidelines are used.

Survey - a uniquely identifiable programme of data collection such as a research cruise, moored instrument deployment or survey event

Project - a collection of surveys that have been completed for a common purpose

B   Detailed Metadata

The detailed metadata are specific to a technique of data collection (e.g. trawl, grab *etc.*) and are subsequently specific to each Data Guideline.

Acquisition Method (Data Production Tools) – details of any method or instruments used to collect the data

C   Data

The tables are specific to the data being collected, and will hold a combination of data and/or metadata required to ensure their re-use.

2.3   Controlled Vocabularies

MEDIN makes use of controlled vocabularies (sometimes called "Term Lists") to ensure that information provided alongside data is unambiguous. The available catalogues of controlled vocabularies used for this MEDIN data guideline are provided primarily by the UK Natural Environment Research Council (NERC), the International Council for the Sea (ICES) and the European Petroleum Survey Group (EPSG). If a term is not available in a recommended list then please contact MEDIN to arrange for the term to be added.

The NERC lists may be viewed at
https://www.bodc.ac.uk/data/codes_and_formats/vocabulary_search/. The search client allows users to interrogate specific lists or to search throughout the entire database for suitable terms. Users can specify how to group the output and may export search results into an MS Excel file.

The ICES term lists are available at http://vocab.ices.dk/
Use the search box to find term lists; users can also select the themes they require to filter their searches. The results are shown for the selected list and may be downloaded into MS Excel by selecting the Excel symbol at the top right of the list.

The EPSG database of coordinate reference systems (http://www.epsg-registry.org/) provides a dictionary of reference systems with a code for each entry. In brief, to find a code, enter the title (e.g. WGS84) into the 'Name' field and click search. The name, code and further information are displayed. If a user is looking for a specific type of reference system such as 'vertical' then click in the 'Type' box, hover over coordinate reference system and click on vertical and then click the search button and all recorded vertical reference systems are shown. If users have a requirement to search for a reference system in a particular part of the world (e.g. Northern Ireland Grid) they may do so by submitting a term to the 'Area' box or fill out the latitudes and longitudes then click search. The website also provides a database of the reference systems and web services to access the information.

2.4   Relationship between MEDIN data guidelines and MEDIN discovery metadata

The MEDIN discovery metadata format is aimed at allowing the non-informed user to discover data sets and it is likely that one 'discovery' data set record will contain a large range of data types that are in turn covered by a range of data guidelines. To enable individuals to reuse data of a specific

nature (e.g. benthic invertebrate data) then related information must be collected (e.g. data owner, reference systems used etc.). Some of the information which is collected in the General Metadata in a data guideline is also required to create a discovery metadata record. Who creates the MEDIN discovery record for a dataset is case specific and dependant on the organisation, and the relationship it has with a Data Archive Centre. However it is intended that the information collected at the 'Survey Information' level is reused for creating a MEDIN discovery metadata record. Further details are available on the MEDIN website which demonstrates clearly which fields in the MEDIN Data Guidelines can be reused for which elements in the MEDIN Discovery Metadata Standard.

2.5   Updates and Feedback

If users have any comments or feedback on these guidelines please contact enquiries@medin.org.uk. Standards develop over time and it is likely that this standard will change in the future. MEDIN advise that users return to the MEDIN website to identify new versions and that they sign up to the MEDIN Standards e-mail listing (e-mail enquiries@medin.org.uk) and Marine Data News to be kept informed of developments.