



www.oceannet.org

Title	MEDIN data guideline for the recording of water sample data.
MEDIN Discipline	Marine Chemistry, Physical Oceanography, Marine Biodiversity, Anthropogenic Properties
Author(s)	M. Charlesworth, G. Evans
Document Owner	M. Charlesworth
Reviewed by	MEDIN Data Standards Group, S. Gontarek, G. Dawson,
Date reviewed	June 2010
Version	3.1
Date approved and published on MEDIN website	18 July 2010
Date last checked for accuracy	18 July 2010
Summary	This guideline defines the format of data and information produced from the collection of discrete water samples which are subsequently used for chemical or biological determination. If used correctly the data will be easily used and reused.
Keywords	Oceanography, water, sample, nutrients, contaminants, chemistry.

Change history		
Version	Date	Change
1.0	22/12/09	First draft of document
3.0	22/05/10	Redrafted in new data guideline structure
3.1	13/07/10	Redrafted following review and published

1.1. Background

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

- Allows contracting organisations 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
- Instills good practice amongst users

Each standard 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 standards can be accessed through www.oceannet.org.

This guideline has drawn heavily on the work of the ICES Working Group on Data Information Management that have produced Guidelines to describe the elements of data and metadata important to the ocean research community. These guidelines are targeted toward physical-chemical-biological data types collected on oceanographic research vessel cruises. Each guideline addresses the data and metadata requirements of a specific data type. These guideline scan be accessed at <http://www.ices.dk/datacentre/guidelines/DataTypeInfoelines/DataTypeInfoelines.s.asp>

This MEDIN guideline sits within the scope of the ICES guideline but differs in that it recommends structures and term lists to be used.

1.2. Scope

This guideline covers the collection of discrete water samples which are then used for chemical or biological determination. This guideline does not cover in-situ measurements. Other more detailed guidelines exist for specific analyses however this guideline provides a general structure and content that should be seen as a minimum. It covers both the raw data from such sampling, methodologies used and derived summary information.

1.3. Using this data guideline

This guideline is split into sections that refers to information that can be collated at different levels. Information that is likely to be the same for all samples (e.g.

ship used, datums used) is collated in the 'Project' and 'Survey' table and the instrument and processing information is held in the Data Production Tools (Sample Methods and Processing Techniques) table. Information that is specific to each profile is collected in the 'Sample Event Table' and the raw profile data is collected in the 'Sample Data' table. If the sampling programme is based upon return visits to a specific station and location then these details should be submitted in the 'Fixed Station Table'.

Some information stated in this guideline is common to all MEDIN guidelines and may be used in part to derive a MEDIN discovery metadata record. Where the survey is part of a ship cruise then the cruise report may hold the required information.

The tables below outline the data fields, a description and where available a term list and/or format given at the end of each field which should be used to store the data. Each field is either mandatory, conditional or optional as indicated by M, C, or O respectively. In the absence of an existing spreadsheet or database to hold the below information, it is recommended that the template available to download from the [MEDIN website](#) is used. Instructions are provided in the template.

To submit this data to a Data Archive Centre the data must be saved and transferred in the .csv file format.

1.4. Further information on the SeaDataNet, ICES and EPSG term lists

The available catalogues of term lists used for this MEDIN data guideline are provided primarily by SeaDataNet, the International Council for the Sea (ICES) and EPSG. If a term is not available in a recommended list then please contact MEDIN to arrange for the term to be added.

The SeaDataNet list may be viewed at http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx. By clicking on the list any term may be searched for by using the drop-down menus or all terms viewed by clicking search. The terms may be viewed in groups of 15 or may be downloaded into an excel file.

The ICES term lists are available at <http://www.ices.dk/datacentre/reco/>. Select which list you require from the 'Reference Code List' drop-down box. The results are shown for the selected list and may be downloaded into MS Excel by selecting the inverted green arrow.

There are a number of ways of describing a spatial dataset. Common horizontal coordinate reference systems include WGS84 and British National Grid. Common vertical coordinate reference systems include Highest Astronomical Tide and Ordnance Datum Newlyn (ODN). It is important that which coordinate

reference system used for a data set is recorded so conversions can be carried out between reference systems. The EPSG database of coordinate reference systems (<http://www.epsg.org/Geodetic.html>) provides a dictionary of reference systems. In brief, to find a code click on the OGP Online Registry and if you know the title (eg WGS84) then type this in the 'Name' field and click search. The name, code and further information is displayed. If you are 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 you want to search for a reference system in a particular part of the world (e.g. Northern Ireland Grid) then you may do so by submitting a term to the 'Area' box or fill out the lat and longs then click search. The website also provides a database of the reference systems and web services to access the information.

1.5. 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 at the Survey Level 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.

2.1. Project Information.

If your collection of data forms part of a wider project or time series then the below details must be recorded.

M, C, O indicate which fields are mandatory, conditional or optional.

Heading	M, C, O	Description	Recommended Term List or Format
Project name	M	The nationally/internationally accepted version of the project name	Free text; (e.g. Rapid Climate Change)
Project website	C	If a Project website exists give the address	e.g. (http://www.noc.soton.ac.uk/rapid/rapid.php)
Project start date	M	The date that the project started	Date; yyyy-mm-dd; (e.g. 2001-01-24)
Project end date	C	The date that the project is due to finish	Date; yyyy-mm-dd; (e.g. 2007-01-24)
Project code	M	Provide a code to uniquely identify the project and allow links to be made between the tables.	Free text; (e.g. RCC)

2.2. Survey Information (Data Activity).

The survey information is a uniquely identifiable programme of data collection such as a research cruise or survey event. This information is likely to be the same for all sample events (e.g. stations) and subsamples in a given data set such as a cruise. Note that in the event that these are not common to all sample events then they should be specified for each one. These fields are common throughout many other MEDIN data guidelines and only need to be given once and referenced if your data set is composed of many data types and therefore conforms to a number of MEDIN Data Guidelines. Where data collection is undertaken on research vessels the data below can often be sourced in the Cruise Summary Report.

M, C, O indicate which fields are mandatory, optional or conditional respectively.

Heading	M, C, O	Description	Recommended Term List or Format
Survey name	M	Title of the survey	Free text; (e.g. Menai Straight Benthic Survey 2004)
Survey description	M	Brief description of the purpose of the survey and other types of measurements that were made for the survey.	Free Text
Survey code	M	A unique code for the survey to allow links to be built between this and sample event data (the cruise identifier code could be used)	Free text; (e.g. JCR3022)
Responsible organisation	M	Organisation who has funded the work	Term List; European Directory of Marine Organisations (e.g. 28: Centre for Environment, Fisheries and Aquaculture Science, Lowestoft Laboratory)
Survey start date	M	The date and time that the survey started.	Date or DateTime; yyyy-mm-dd or yyyy-mm-dd hh:mm:ss (e.g. 2009-01-24 12:33:00)
Survey end date	M	The date and time that the survey ended.	Date or DateTime; yyyy-mm-dd or yyyy-mm-dd hh:mm:ss (e.g. 2009-02-16 16:33:00)
Spatial coordinate reference system	M	Describes the system of spatial referencing. I.e. the datum used to provide details of latitude and longitude.	Term List; http://www.epsg.org/Geodetic.html (e.g. WGS84 is EPSG::7030)
Position fix	M	Give the method and source of the position	Free Text; (e.g. Differential GPS taken from

method and source		fix instrument.	the ships navigation equipment.
Horizontal positional accuracy	M	How accurate the spatial positions are likely to be	Number; units = meters (e.g. 15)
Depth coordinate reference system	C	Give the reference to which the depth has been calculated e.g. Highest astronomical tide. Mandatory if seabed depths are given for each sample.	Term List http://www.epsg.org/Geodetic.html (e.g. ODN is EPSG::5701)
Vertical positional accuracy	C	How accurate the vertical resolution is. Must be provided if seabed depths are given.	Number; units = meters (e.g. 0.5)
Platform type	O	The platform type (e.g. Research Vessel) from which the sampling device was deployed.	Term list SeadataNet Platform Classes (L061) (e.g. 31)
Ship name	O	The name of the ship from which the sampling device was deployed.	Term list SHIPC at http://www.ices.dk/datacentre/reco/ (e.g. 74LG Lough Foyle)
Cruise report reference	O	Cruise report reference if applicable.	Free text; in reference format. e.g. Litt, E.J. 2009. PHiXT 4. 30 July to 2 August 2009 <i>RV Prince Madog</i> POL Coastal Observatory Liverpool Bay Cruise Report. POL Coastal Observatory, Liverpool.
Project code	M	State the code of the project given in the project table to allow links to be made between the tables.	Free text; (e.g. RCC)

2.3. Fixed Station Information.

In many cases a fixed point, transect or area is returned to on a number of occasions to form a time series. The actual sample event may not be in exactly the same location each time however due to ship movements or sampling strategy, however it is useful to record both the position which is intended to be sampled (fixed) and the actual sampling position (sample).

Therefore, the information below must be included if a fixed point, transect or area is used as the basis for replicate profiles or for repeat monitoring surveys. Actual profile coordinates should be placed in the sample event table. A fixed station may be a point, transect, or an area. If the fixed station is a transect or an area then the secondary latitude and longitude fields must be completed.

Heading	M, C, O	Description	Recommended Term List or Format
Local station identifier	M	A unique identifier for the station	Free text. e.g. Stanton Bank site 4 (point) e.g. Liverpool/Dublin transect (transect) e.g. Lagan Estuary (area)
Primary Latitude (decimal degrees)	M	The primary latitude of the fixed station given in decimal degrees. For a point this field is set to the point latitude; for a transect it is set to the latitude of the start of the transect; for an area it is set to the southern edge of the box. Units are positive north.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. 54.5837
Primary Longitude (decimal degrees)	M	The primary longitude of the sample given in decimal degrees. For a point this field is set to the point longitude; for a transect it is set to the longitude of the start of the transect; for an area it is set to the western edge of the box. Units are positive east (West is negative, East is positive).	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. -5.5837
Secondary Latitude (decimal degrees)	C	The secondary latitude of the fixed station given in decimal degrees. For a point this field is not required; for a transect it is set to the latitude of the end of the transect; for an area it is set to the northern edge of the box. Units are positive north.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. 55.7393

Secondary Longitude (decimal degrees)	C	The secondary longitude of the sample given in decimal degrees. For a point this field is not required; for a transect it is set to the longitude of the end of the transect; for an area it is set to the eastern edge of the box. Units are positive east (West is negative, East is positive).	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. -3.7394
Description of fixed station spatial form	M	Describe if the fixed station is a point, transect or an area.	Term list; <u>SeadataNet Geospatial Feature Type (L021)</u> (e.g. point)

2.4. Sample Event (Water Sample). This table holds information on the location and time of each sample. In most cases a number of water samples are taken at a number of depths at each station and therefore many fields in this table will be consistent across those water samples.

Heading	M, C, O	Description	Recommended Term List or Format
Survey code	M	Give the survey code to allow links to be built between this table and survey data table.	Free text; (e.g. JCR3022)
Local station identifier	M	A unique identifier for the station under consideration	e.g. Stanton Bank site 4 e.g. PS74926
Latitude of sample given in original recorded format	M	The latitude of the sample given in whichever format was used to record at the time of sampling. Units are positive north.	e.g. 50°47'24"
Longitude of sample given in original recorded format	M	The longitude of the sample given in whichever format was used to record at the time of sampling. Units are positive east.	e.g. -4°21'53"
Latitude of sample (decimal degrees)	M	The latitude of the sample given in decimal degrees. Units are positive north.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. 54.5837
Longitude of sample (decimal degrees)	M	The longitude of the sample given in decimal degrees. Units are positive east.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. -3.476
Time zone	M	Give the time zone in which the date and time of the sample collection is made (preferably Coordinated Universal Time (UTC))	Free Text; (e.g. UTC)
Date and time	M	The date and time of sample collection.	yyyy-mm-dd or yyyy-mm-dd hh:mm:ss (e.g. 2009-01-24 13:33:00)
Sample	M	A unique identifier for the sample under	Text; (e.g. E5; PHJ7936)

Identifier		consideration	
Rosette position	C	If the sample was taken from a rosette then give the position of the bottle on the rosette from where the sample was taken	Number (e.g. 5)
Sample Depth	M	Give the depth at which the sample was taken	Number; units = meters (e.g. 150)
Sampling personnel	O	Names or the personnel who were involved in collecting and field processing the samples	Text, personnel initials given and separated by semi-colon if more than one personnel used; (e.g. J. Bloggs; B. Begger collected and field processed samples)
Sample collection notes	O	Any further notes on the sample collection that may be of relevance	Text; (e.g. Due to rough weather the rosette was not stable when it reached the surface).
Water Column Depth	M	The total water column depth	Number; units = meters (e.g. 1500)
Associated Information	C	If subsamples were taken from the same sample and used for different analysis give details here (e.g. chlorophyll, nutrient, metal).	Text; (e.g. Dissolved Oxygen analysis was carried out as a separate sample at each depth)
Link to CTD profile	C	If a CTD profile was taken at the same time then give the reference number of that profile to allow links between the sample and physical measurements to be made	e.g. CTD009 e.g. 69729

2.5. Sample Analysis and Techniques (Data Production Tools). If the sample analysis and techniques are different for each sampling event then the fields in this table should be appended to the sample data and be completed for each sample. In many cases the information in this table is consistent for a whole survey in which case it should only have to be completed once. Where necessary the information in this table should be completed for each parameter under consideration. Information in this table may also be used to complete fields in the discovery metadata.

Heading	M, O, C	Description	Recommended Term List or Format
Survey code	M	Provide the survey code to allow links to be built between the analysis and techniques for each parameter and the survey. If analysis and techniques are not consistent for a survey then this information should be linked to each sample data result.	Free text; (e.g. Menai Straight Benthic Survey 2004)
Sampler Instrument Details	M	Give sampling instrument description, manufacturer and model if applicable	Term Code list SMTYP at http://www.ices.dk/datacentre/reco/ (e.g. NSK, Niskin Bottle)
Sample Processing Details	M	Give a description of any sample processing conducted before the sample was presented for analysis or stored. In particular give details of any filtration	Free text; (e.g. Niskin bottles were removed from the rosette and the contents split between 3 subsamples. The first subsample for nutrient analysis was frozen immediately; the second was filtered through a GFF filter and then 10ml of 90% acetone analysis added; the 3 rd subsample was fixed with 3 drops of Lugol's Iodine and kept for phytoplankton analyses.)
Analytical Details	C	Give details of the analysis of the water samples	Free text; (e.g. Nutrients were measured by nitrate reduction following the method of Strickland (1972) at CEFAS Lowestoft; e.g. phytoplankton abundance was obtained by settling the fixed samples in settling chambers and examining them under an inverted microscope)

Data Processing Details	C	If the data has been processed then detail the steps here including, de-spiking or smoothing methods, editing and quality control methods, and an overview report.	Free text; (e.g. Data was checked to ensure it was within range for the time of year and location and there were no outliers. Any suspected outliers were flagged)
Analytical Laboratory or Organisation	M	The laboratory/organisation that analysed the samples	Reference Code list RLABO at http://www.ices.dk/datacentre/reco/ (e.g. UNIC Unicomarine Ltd, Letchworth Laboratory)
Analytical personnel	O	Names or the personnel who were involved in analysing the samples	Text, personnel initials given and separated by semi-colon if more than one personnel used; (e.g. J. Bloggs analysed all samples).
Analytical notes	O	Any further notes on sample analysis that may be of relevance	Text; (e.g. Each batch of 10 nutrient samples were run with a standard and blank)
Quality Control Scheme	M	Description of any quality control scheme that the laboratory participated in during the analysis.	Text; (e.g. National Marine Chemistry Analytical Quality Control Scheme)

2.6. Sample Data. When providing the sample data it must be clearly linked to the station and sample event information. The preferred format for holding the raw data is in where each record is presented as a row. However it is recognized that often results are produced in a matrix format and this may be used for specific applications.

Heading	M, O, C	Description	Recommended Term List or Format
Sample Identifier	M	The unique identifier for the sample under consideration as detailed in the sample event table.	Text; (e.g. E5; PHJ7936)
Parameter (chemical) or species (biological)	M	Give parameter or species name where possible or higher taxonomic group if not.	For species such as plankton use reference code list http://www.marinespecies.org/ (e.g. <i>Fragilariforma virescens</i>) Reference code list P011 at http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx (e.g. CHEMM012, Concentration of nitrate {NO ₃ } per unit volume of the water body [dissolved plus reactive particulate phase] by colorimetric autoanalysis and corrected for nitrite)
Value (abundance)	M	For chemical determinants give the value of the measurement to the number of decimal places that can be given based on the accuracy of the analysis. For biological species give the abundance	For chemical determinands give number; (e.g. 0.552). For biological if counts of individual taxon then give number (e.g. 34). If presence/absence then give P/A.
Unit of measurement	M	Give the unit of measurement	Reference Code list P061 at http://seadatanet.maris2.nl/v_bodc_vocab/welcome.aspx (e.g. UGPL, ug/l)
Biomass (species only)	O	Biomass of individuals in 1 taxon group	Integer Number units = g. (e.g. 0.23)

Stage development of taxon (species only)	<input type="radio"/>	Development stage of the taxa	Reference Code list STAGE at http://www.ices.dk/datacentre/reco/ (e.g. JU Juvenile)
Photographs and videos (species only)	<input type="radio"/>	Note if images were taken of specimens at any stage of the processing, the purpose they were collected for, where they are held, what their IDs are and what format the images are in.	Text; (e.g. Images taken of <i>Mytilus edulis</i> were taken to confirm if it was a new subspecies. Images submitted to MEDIN using data guideline on digital images. Images reference numbers are: Mytilus_02mar08_01 to Mytilus_02mar08_68)
Specimens (species only)	<input type="radio"/>	If specimens of the taxa are held then indicate here	Text; (e.g. Examples of all the species within the family Mollusca recorded are stored)
Notes	<input type="radio"/>	Any further notes that may be of relevance	Text (e.g. The template provided in the MEDIN data guideline was used)