www.oceannet.org

Title	MEDIN data guideline for the recording of surface underway oceanographic data.
MEDIN DIscipline	Physical Oceanography
Author(s)	M. Charlesworth,
Document Owner	M. Charlesworth
Reviewed by	MEDIN Data Standards Group; M. Hearn
Date reviewed	July 2010
Version	3.0
Date approved and published on MEDIN website	15 July 2010
Date last checked for accuracy	15 July 2010
Summary	This guideline defines the format of data and information produced from the collection of surface oceanographic data from instruments fixed on a vessel while underway. If used correctly the data will be easily used and reused. A xls template is provided if required.
Keywords	CTD, Oceanography, Underway, Salinity, Conductivity, Temperature, Depth, DO, turbidity, fluorescence.

Change history		
Version	**Date**	**Change**
1.0	12/01/10	First draft of document
2.0	27/05/10	Redraft to take into account new structure and comments on the ocean profile data guideline.
3.0	14/07/2010	Minor edits following comments from reviewers.

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 organisation 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 has 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/DataTypeGuidelines/DataTypeGuidelines.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 defines the format of data and information produced from the collection of oceanographic data from moored instruments. It covers both the raw data, 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 or transfer to other organisations the raw data should be provided in the file type outputted from the instrument. Other tables should be provided in the .csv 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) the 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 *RV Prince Madog* 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; SeadataNet Geospatial Feature Type (L021) (e.g. point)

2.4. Sample Event (Transect Information). This table holds information on the location and time of the sampling transect. In some instances a number of transects may be deployed on the same survey or cruise and therefore the information in this table will be required for each one.

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)
Transect identifier	M	A unique identifier for the transect under consideration	e.g. NW Irish Sea north/south e.g. PS74926
Transect Description	O	Provide a brief description of the transect	Free text; (e.g. transect made at ½ degree intervals of latitude through the NW Irish Sea)
Start Date and time	M	The start date and time of the transect.	yyyy-mm-dd or yyyy-mm-dd hh:mm:ss (e.g. 2009-01-24 13:33:00)
Start Latitude of sample given in original recorded format	M	The start latitude of the transect given in whichever format was used to record at the time of sampling. Units are positive north.	e.g. 50°47’24’’
Start Longitude of sample given in original recorded format	M	The start longitude of the transect given in whichever format was used to record at the time of sampling. Units are positive east.	e.g. -4°21’53’’
Start Latitude of sample (decimal degrees)	M	The start latitude of the transect given in decimal degrees. Units are positive north.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. 54.5837
Start Longitude of sample	M	The start longitude of the transect given in decimal degrees. Units are positive east.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. -3.476

(decimal degrees)			
End Date and time	M	The end date and time of the transect.	yyyy-mm-dd or yyyy-mm-dd hh:mm:ss (e.g. 2009-01-24 13:33:00)
End Latitude of sample given in original recorded format	M	The end latitude of the transect given in whichever format was used to record at the time of sampling. Units are positive north.	e.g. 50°47’24’’
End Longitude of sample given in original recorded format	M	The end longitude of the transect given in whichever format was used to record at the time of sampling. Units are positive east.	e.g. -4°21’53’’
End Latitude of sample (decimal degrees)	M	The end latitude of the transect given in decimal degrees. Units are positive north.	Decimal degrees; minimum of two and a maximum of five decimal places. e.g. 54.5837
End Longitude of sample (decimal degrees)	M	The end longitude of the transect 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)
Water bottle sample details	C	If water samples were also taken during the transect then details should be given here to allow the water sample data to be linked to the transect data	Free text; Water samples were taken at 10, minute intervals and determined for nutrients, and chlorophyll concentrations. Data is recorded in the Bottle database under the same station

			and profile identifiers as recorded here.

2.5. Sample Methods, Instruments and Processing Techniques (Data Production Tools). If the information in this category is the same for all transects within a survey then the Survey Name must be completed to allow links to the survey information. The instrument details should be provided for each instrument sensor and a clear relationship made to the sample data file headings to allow a data manager to determine which channels (or headings) within the files refer to which instrument sensors.

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)
Depth of water intake	M	Give the depth at which the water intake is.	Number; units = meters (e.g. 0.5)
Flow details	M	Describe the length and diameter of tube between intake and sensors and the flow rate.	Free text; (e.g. a 10mm diameter polyethylene tube was 12 meters in length between intake and instruments and had a flow rate of 50 l/min)
Instrument Details	M	Instrument description, reference number, manufacturer and model — provide a literature reference, web site reference or briefly describe. Include — accuracy, resolution and response range of individual sensors.	Free text; (e.g. SeaBird CTD model number 3756a, serial number BX472946HJK647, accuracy, resolution and response range of sensors can be viewed at www.seabird.com. Flourometer was a Valeport FL510 serial number 53292735).
Instrument Sensor to Data Link	M	Provide details of which instrument sensors apply to which data file headings.	Free text; (e.g. The conductivity, temperature and depth sensors on instrument BX472946HJK647 refer to the field headings CON, TEMP, DEP respectively in the sample data files SB_100m_001 to SB_100m_034)
Instrument Calibration	C	If calibrations have been applied to the instrument a description should be given here including the date of the calibration.	Free text; (e.g. The CTD and Flourometer were returned to Seabird on the 24/03/2007 and calibrated according to their standards in the laboratory)
Instrument	O	Give the software and operating system used to	Free text; (e.g.Seabird software used to

data retrieval		retrieve and record the information from the instruments	retrieve binary file which was then recorded using MS Excel and manipulated using the Windows 2005 operating system).
Water Sample Calibration Details	C	If water samples have been taken to calibrate the instrument then details of those should be given here including any field and laboratory coefficients used. These should include description of or reference to full laboratory methods and procedures, Details of any external sample analysis, including the laboratory name and accreditation level, A description of or reference to any internal or external quality assurance procedures.	Free text; (e.g. Water samples were taken at regular depths and chlorophyll measured to calibrate the flourometer following the method of Bloggs (1972) at CEFAS Lowestoft laboratory which participates in the UK National Marine Chemistry Analytical QC scheme).
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. Sampling intervals and any corrections for time lag between stations should also be recorded.	Free text; (e.g. All instrument data was screened and despiked following expert examination. Any data that was suspected to be incorrect was flagged, etc etc)

2.6. Sample Data. Surface underway data typically consists of a combination of position, time, pressure, conductivity, temperature and any other parameters measured such as turbidity. It is recommended that the raw files from the instrument are supplied as well as any additional files to which calibrations may have been applied to. These data should be provided in a readable form with the header information sufficiently explained. The recommended format for surface underway data is as a matrix with date/time as the primary channel and other parameters details as additional fields. The transect name should be identified either in the file names or within the file to allow the data to be matched to a transect within a survey. Please ensure that there is a clear relationship between the instrument sensors calibration information and the fields (channels) for each sensor within the sample data files.