

# ICES Guidelines For Discrete Water Sample Data

(Compiled December 1999, revised August 2001; June 2006)

In the context of this guideline, discrete water sample data are considered to be any data that result from a single collection of water and so covers a huge variety of parameters. This collection of water must have a specific, identifiable time, position and depth. Such data could originate from a single bottle attached to a rosette or water drawn from a non-toxic supply.

No integrated samples are considered as part of discrete water sample data. Thus, tows that result in integrated data values are not considered in discrete water sample data. Nor are integrated samples from a pumping system or sediment trap.

## 1.0 Receiving Data

The Data Centres require the following information to be supplied by the data supplier together with the data. When receiving data, the Data Centres of the ICES community shall strive to meet the following guidelines.

### 1.1 *Data standard*

All parameters must be clearly specified and described. If parameter codes are to be used, then the source data dictionary consistency must be specified. Parameter units must be clearly stated. Parameter scales must be noted where applicable. If computed values are included, the equations used in the computations should be stated.

The data should be fully checked for quality and pre-edited or flagged for erroneous values. An explicit statement should be made of the checks and edits applied to the data.

A brief description, or a reference to the data collection and processing methods (e.g. reference to a specific technique or specific project protocols) must be included and should contain information regarding:

- Describe or reference full laboratory methods and procedures
- If sample was sent out for analysis, give laboratory name and accreditation level
- Describe or reference any internal or external quality assurance procedures (e.g. QUASIMEME, IAPSO)

A brief description of the data processing procedures must be included and should contain information regarding:

- editing/quality control methods
- how are trace values (values below the detection limit) identified
- how are missing values handled (null vs. zero, or “blanks”)
- what is the precision of the methods (e.g. number of significant figures)

- what analyses has been performed (use parameters descriptions as described in the ICES green book)
- what units are used
- whether any duplicate samples were taken
- describe what quality flags are used if any
- comments describing each station
- supply a calibration document

If a cruise/data report is available describing the data collection and processing, this can be referenced. If possible a copy should be supplied with the data.

## **1.2 Format description**

Data should be supplied in a fully documented ASCII format. Data Centres are capable of handling water sample data in a wide variety of user-defined and project formats. If in doubt about the suitability of any particular format, advice from the Data Centre should be sought.

Individual fields, units, etc. should be clearly defined and time zone stated. Time reported in UTC is strongly recommended. Ideally all of the data from the single water source should be stored in a single file. The contents of the data and ancillary information should adhere to the Formatting Guidelines for Oceanographic Data Exchange ([http://ocean.ices.dk/formats/GETADE\\_Guidelines.aspx](http://ocean.ices.dk/formats/GETADE_Guidelines.aspx)) prepared by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) and available from RNODC Formats.

Often different groups or laboratories will analyse a single water sample for a multitude of parameters. In such cases, it is common for the data from the different groups to arrive at the data centre at different times. The receiving data centre may merge those data from a single water source. Thus it is crucial that the date/time, position and sample identifier accompany the data.

## **1.3 Collection Details**

Pertinent information to be included in the data transfer to the Data Centre includes:

- Project, platform, cruise identifier
- Country, organisation, institute, PI
- Station number, site details, sample identifier (or bottle number), , type of station (CTD, CTD(NMMP), continuous flow etc.,
- Analyses performed e.g. salinity and nutrients
- Date and time of the start of the sampling and date of analysis (UTC is recommended)
- Position (latitude and longitude degrees and minutes or decimal degrees can be used. Explicitly state which format is being used. It is recommended that N, S, E and W labels are used instead of plus and minus signs.)
- Description of operational procedures including (where applicable) sampling rate,

detection limits, standard analytic procedures, calibration of equipment, quality control of original data, methods of position fixing (e.g. GPS, DGPS)

- Details of the collection instrument and sensor (e.g. manufacturer, model number, serial number, and sampling rate)
- Sounding (station depth and sample depth) should be included for each station. The method and assumptions of determining the sounding should be included.
- Type of analyses undertaken including any nutrient samples analysed
- Range of data values (desirable)

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use.

For additional information on quality control procedures, metadata requirements for particular parameters and collection instrumentation, see UNESCO (1996).

## **2.0 Value Added Service**

When processing and quality controlling data, the Data Centres of the ICES community shall strive to meet the following guidelines.

### **2.1 Quality Control**

A range of checks are carried out on the data to ensure that they have been imported into the Data Centre's format correctly and without any loss of information. For discrete water sample data, these should include:

- Check header details (vessel, cruise number, station numbers, date/time, latitude/longitude (start and end), instrument number and type, station depth, cast (up and down) data type /no. of data points, platform identifier)
- Plot station positions to check not on land
- Check ship speed between stations to look for incorrect position or date/time
- Automatic range checking of each parameter (e.g. WOD 1998, Maillard 2000)
- Check units of parameters supplied
- Check pressure increasing or decreasing as appropriate
- Check no data points below bottom depth
- Check depths against echo sounder
- Plot profiles (individually, in groups, etc)
- Check for spikes
- Check for vertical stability/inversions
- Check profiles vs. regional climatology
- Check calibration information available
- Compare parameters for predictable relationships (e.g. parameter ratios)
- Check for consecutive constant values

- Duplicate detection when comparing to archived data
- Flag suspicious data or correct after consultation with Principal Investigator (PI)

## **2.2 Problem Resolution**

The quality control procedures followed by the Data Centres will typically identify problems with the data and/or metadata. The Data Centre will resolve these problems through consultation with the originating PI or data supplier. Other experts in the field or other Data Centres may also be consulted.

## **2.3 History Documentation**

All quality control procedures applied to a dataset are fully documented by the Data Centre. As well, all quality control applied to a dataset should accompany that dataset. All problems and resulting resolutions will also be documented with the aim to help all parties involved; the Collectors, Data Centre, and Users. A history record will be produced detailing any data changes (including dates of the changes) that the Data Centre may make.

## **3.0 Request for Support**

When addressing a request for information and/or data from the User Community, the Data Centres of the ICES community shall strive to provide well-defined data and products. To meet this objective, the Data Centres will follow these guidelines.

### **3.1 Data Description**

The Data Centre shall aim to provide to its clients well-defined data or products. If digital data are provided, the Data Centre will provide sufficient self-explanatory series header information and documentation to accompany the data so that they are adequately qualified and can be used with confidence by scientists/engineers other than those responsible for their original collection, processing and quality control. This is described in more detail below:

- A data format description fully detailing the format in which the data will be supplied
- Parameter and unit definitions, and scales of reference
- Definition of additional quality control
- Flagging scheme, if flags are used
- Data history document (as described in 3.2 below)
- Accompanying data (e.g. CTD data at the time of bottle trip)

### **3.2 Data History**

A data history document will be supplied with the data to include the following:

- A description of data collection and processing procedures as supplied by the data collector (as specified in Section 1.1 and 1.3)
- Quality control procedures used to check the data (as specified in Section 2.1)
- Any problems encountered with the data and their resolution and modification date
- Any changes made to the data and dates of these changes

Any additional information of use to secondary users which may have affected the data or have a bearing on its subsequent use should also be included.

### **3.3 Referral Service**

ICES member research and operational data centres produce a variety of data analysis products and referral services. By dividing ocean areas into regions of responsibility, and by developing mutually agreed guidelines on the format, data quality and content of the products, better coverage is obtained. By having the scientific experts work in ocean areas with which they are familiar, the necessary local knowledge finds its way into the products. Data and information products are disseminated as widely as possible and via a number of media including mail, electronic mail and bulletin boards.

If the Data Centre is unable to fulfil the client's needs, it will endeavour to provide the client with the name of an organisation and/or person who may be able to assist. In particular, assistance from the network of Data Centres within the ICES Community will be sought.

## **REFERENCES**

Maillard, C. And M. Fichaut. 2000. Medar-Medatlas Protocol, Part I : Exchange Format And Quality Checks For Observed Profiles, IFREMER, June 2000 - R.INT.TMSI/IDM/SISMER/SIS00-084.

UNESCO. 1996. IOC-EU-BSH-NOAA-(WDC-A). International Workshop on Oceanographic Biological and Chemical Data Management Hamburg, Germany 20-23 May 1996, IOC Workshop Report 122.

WOD, 1998. World Ocean Database, Documentation and Quality Control, Version 2, Silver Spring, MD, December 1999.