



Baltic Marine Environment Protection Commission

Making the HELCOM eutrophication assessment
operational (EUTRO-OPER)
Video Meeting, 8 September 2015

EUTRO-OPER 5-2015

Document title	QA/QC guidelines for Ferrybox data (draft)
Code	3-2
Category	CMNT
Agenda Item	3 – Progress of EUTRO-OPER during work phase 4
Submission date	24.8.2015
Submitted by	Estonia, Finland, ICES, Secretariat
Reference	EUTRO-OPER 4-2015

Background

The EUTRO-OPER project has been tasked to prepare guidelines for QA/QC management of the assessment data, to be included as part of the 'EUTRO-OPER manual'. EUTRO-OPER 4-2015 agreed, that while the guidelines for discrete water sample data, prepared by ICES, are already operational, guidelines for the new data types should be prepared.

Action required

The meeting is invited to review the draft and guide the work forward.

Guidelines for information from automatic flow-through systems (Ferrybox)

In this guideline, Ferrybox information is considered to be information derived from automatic flow-through systems implemented on board ships of opportunity. The observations are either original or in validated form, and may be aggregated temporally and spatially to a specific level.

The data must include a distinct time and position. In the case of aggregated information, these may be estimates.

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 shall strive to meet the following guidelines.

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. 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:

- Methods and procedures applied to the analysis of raw data
- Methods / protocols and dataset(s) used for validation, or refer to their original source
- Description or reference any internal or external quality assurance procedures

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

- editing/quality control methods
- how are missing values handled (recommended as “blanks”)
- what is the precision of the methods (e.g. number of significant figures) green book)
- what units are used
- describe what quality flags are used if any
- supply a validation document

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

Format description

Data should be supplied in a fully documented ASCII format. 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. 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) prepared by the IOC's Group of Experts on the Technical Aspects of Data Exchange (GETADE) and available from RNODC Formats.

Collection and processing details

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

- Processing responsible: country, organisation, institute, PI
- Description of flow-through system and measuring instruments / sensors
- Measured parameters
- Products derived through validation procedures
- Resolution of original data
- Details of validation data
- Conversions used for deriving chlorophyll *a* concentration from chlorophyll *a* fluorescence data
- Level of temporal and spatial aggregation used
 - spatial: either raw data, HELCOM 20 km grid or HELCOM assessment area
 - temporal: either raw data or daily / annual assessment period
- Uncertainties on product estimates
- Date and time of the start and end of the sampling (UTC is recommended)
- Position estimate (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 procedure for checking spikes

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).

Value added service

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

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), description of automatic sequence water sampler
- Plot station positions to check not on land
- Automatic range checking of each parameter (e.g. WOD 1998, Maillard 2000)
- Check units of parameters supplied
- Flag suspicious data or correct after consultation with Principal Investigator (PI)

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.

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.

Request for support

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

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 below)
- Accompanying data

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 / data provider (as specified earlier)
- Quality control procedures used to check the data (as specified earlier)
- 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.

References