



Theme 2: Hazardous substances



This is a deliverable from the BalticBOOST project that was coordinated by HELCOM and co-financed by the European Union in 2015-2016 as part of the programme DG ENV/MSFD Action Plans/2016.

WP 2.2 Deliverable 1: Improved data labelling and data flow for hazardous substance

Partners: International Council for the Exploration of the Sea (ICES), Baltic Marine Environment Protection Commission (HELCOM)

Author: ICES, HELCOM

Introduction

The HELCOM core indicators forms the basis for the hazardous substances assessment in the upcoming 'State of the Baltic Sea' report (HOLAS II). BalticBOOST WP 2.2 has developed the needed data-arrangements for the HELCOM hazardous substances assessment system. The data-flow for the core indicators has become operational. Within the frame of the project the development has even been advanced further than first anticipated by implementing the first components needed to calculate the actual core indicators using an R-script instead of only pre-processing the data. The project has also specified the data extraction needs to be used for indicator calculations.

Operationalization of the data-flow for the hazardous substances core indicators build on data submitted to the HELCOM COMBINE database. The WP 2.2. Lead Partner (ICES) has developed a data-flow model that is used to implement the assessment system (Figure 1).

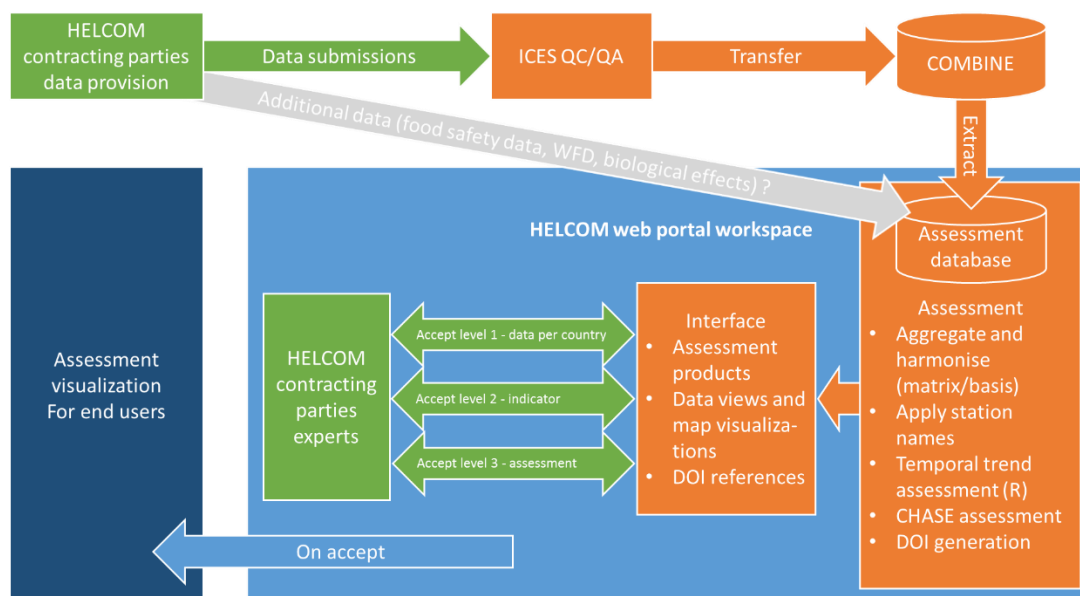


Figure 1. Hazardous substances assessment system data-model. Green boxes identify steps taken by the countries, orange boxes and arrows identify steps taken by the data host ICES and blue boxes identify steps carried out by HELCOM.

The operational hazardous substances assessment system centres around a HELCOM workspace (blue box Figure 1). The workspace allows for reviewing the datasets and indicator evaluations and documentation of all steps and corrections to datasets and indicator calculations etc.

During the lifetime of the project the first components of the assessment system have been implemented in the workspace. The data views to the accessions of data to COMBINE has been implemented (Figure 2).

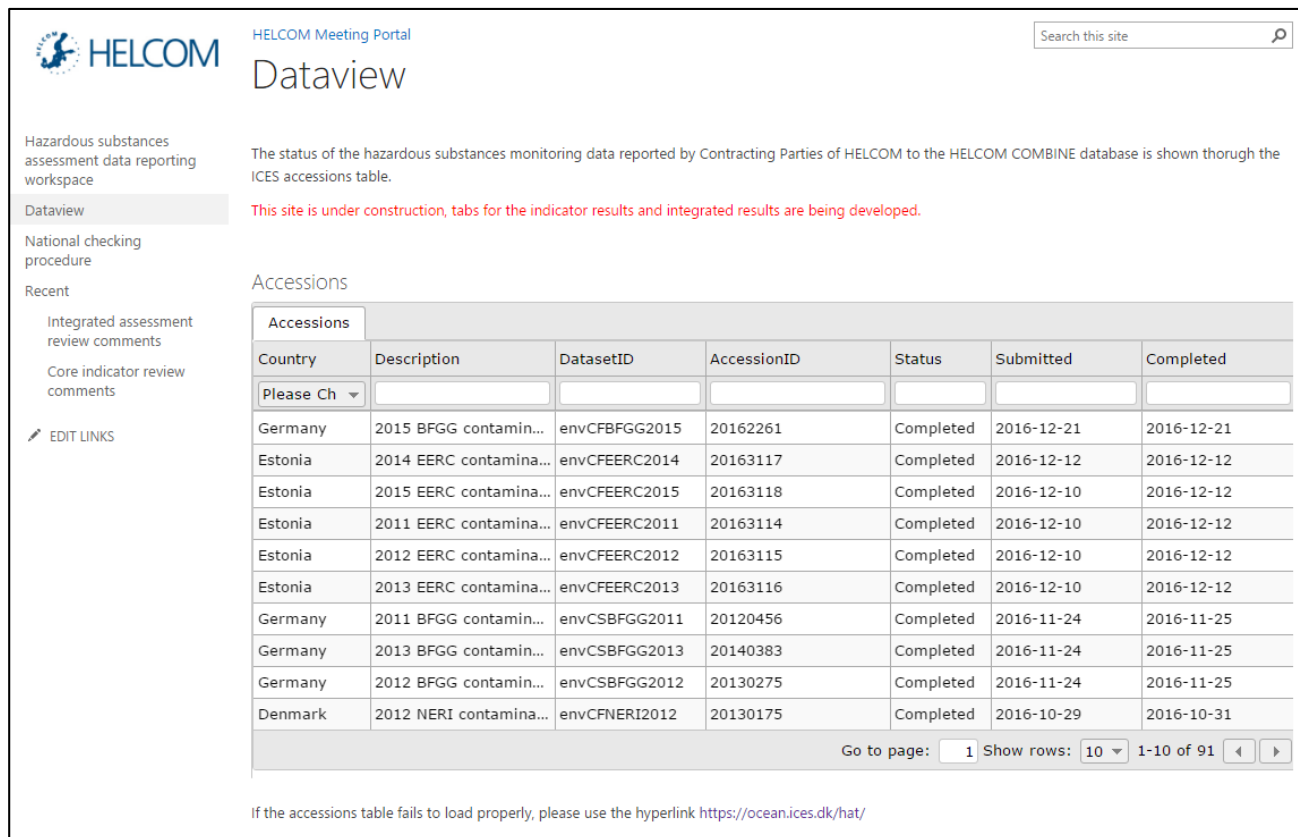


Figure 2. Screen capture of the HELCOM workspace where the data submitted to the HELCOM COMBINE database hosted by ICES by the contracting parties can be seen in real time to follow up on the accessions of the submission.

The development of a spatial overview of the datapoints has also been developed. Currently the spatial overview is available as a separate online widget (<http://gis.ices.dk/sf/index.html?widget=boost>) (Figure 3). The spatial overview of the data will in the future become integrated in the assessment system in the workspace. The spatial overview displays the HELCOM assessment units in the background and also the extent of the territorial waters. This enables the user to review the spatial data distribution with the HELCOM assessment units.

The operational hazardous substances assessment system requires the application of a defined extraction table to the COMBINE database in order to develop the dataset to be used in an assessment, i.e. frozen as an assessment database (Figure 1, orange arrow COMBINE->Assessment database). Through BalticBOOST the extraction table has been developed together with the HELCOM Expert Network on Hazardous Substances (EN-HZ) (Table 1.). The table defines the parameters and required metadata of the parameters to be extracted in order to calculate the HELCOM core indicators.

Procedures for the spatial labelling of data with the HELCOM assessment units, has been implemented for all extracted data including monitoring stations. Progress is currently going on to ensure the option of HELCOM spatial unit labelling in all relevant views published by the ICES Data Centre.

In a test run, the dataset to be used for indicator evaluations in HOLAS II was extracted from the COMBINE database using the developed extraction table (Table 1) and displayed spatially using the online widget (Figure 3).

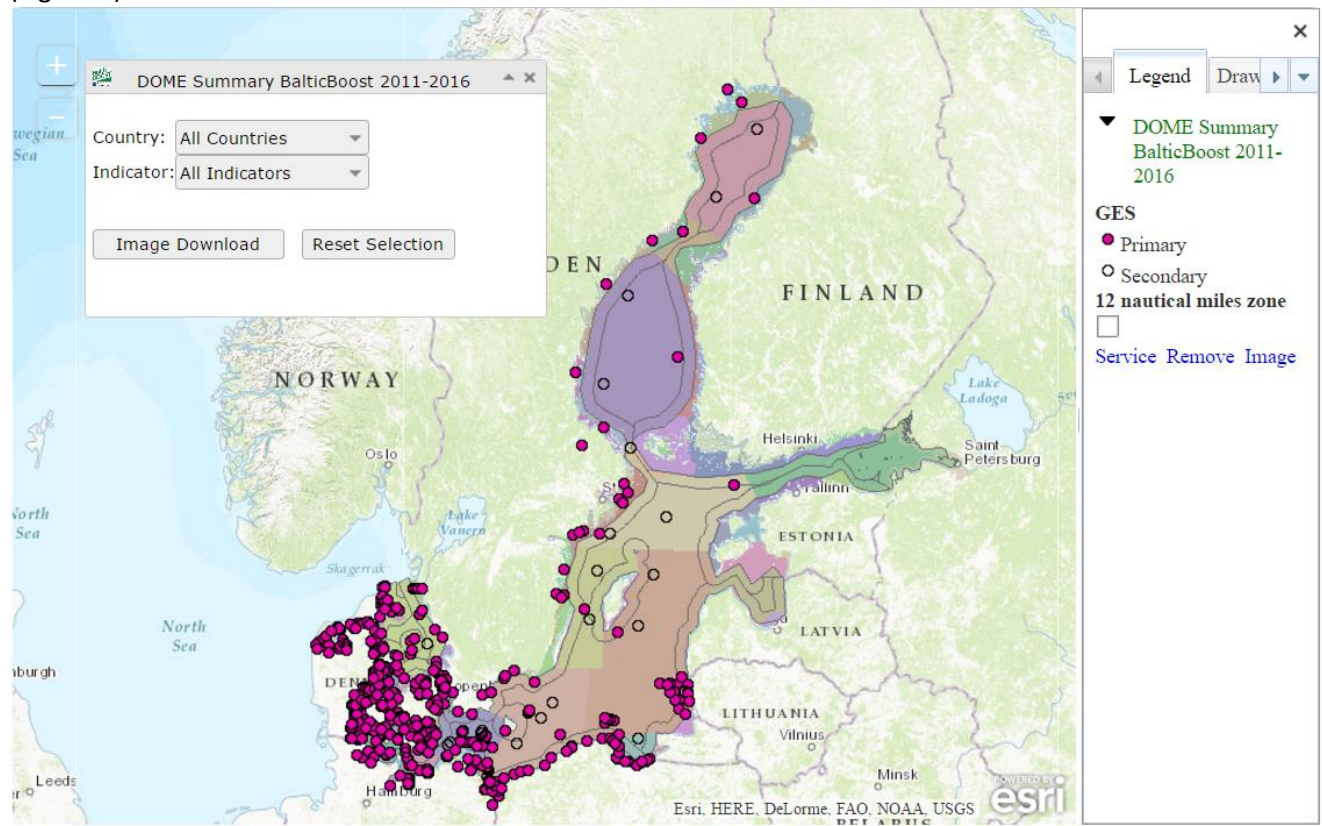


Figure 3. Screen capture of the online widget displaying the assessment data set spatially.

Table 1. Overview table of the parameters, matrices and basis selected for extraction from the COMBINE database to evaluate the core indicators.

NB: For sediment measurements lithium (LI), aluminium (AL), organic carbon content (CORG) and grain sizes should be included in addition to “main” parameters for normalisation

Indicator	Parameters (PARAM) / Parameter groups (PARGROUP) (see also http://vocab.ices.dk/)	Primary matrix / GES	Species	Matrix x	Basis s	Secondary matrix / GES	Species	Matrix ix	Basis s
Metals (Cd) ¹	PARAM = CD	Water				Biota Sediment	Molluscs (M edulis + M. baltica)	SB	W D
Metals (Pb)	PARAM = PB	Water				Biota Sediment	Herring & cod (open sea) Flounder & Perch (coastal) Molluscs (M edulis + M. baltica)	LI SB	W W D
Metals (Hg)	PARAM = HG	Biota	Herring & cod (open sea) Flounder & Perch (coastal) Molluscs (M edulis + M. baltica)	MU (‘fillet) SB	W W				
HBCDD	PARAM = HBCD, (HBCDA,HBCD B, HBCDG)	Biota	Herring & cod (open sea) Flounder & Perch (coastal)	MU (‘fillet) & LI	W	Sediment		All	

¹ Denmark and Estonia have a study reservation on the secondary threshold OSPAR BAC 26 µg/kg ww fish liver. The row is not included in the extraction table.

Indicator	Parameters (PARAM) / Parameter groups (PARGROUP) (see also http://vocab.ices.dk/)	Primary matrix / GES	Species	Matrix x	Basis s	Secondary matrix / GES	Species	Matrix ix	Basis s
PBDE	PARAM = BD28, BD47, BD99, BD100, BD153, BD154	Biota	Herring & cod (open sea) Flounder & Perch (coastal)	MU (‘fillet) & LI	W				
PFOS	PARAM = PFOS	Biota	Herring & cod (open sea) Flounder & Perch (coastal)	MU (‘fillet) & LI	W	Water		WT	
PCB and dioxins and furans ²	Dioxins and furans: PARGROUP = OC-DX	Biota	Herring & cod (open sea) Flounder & Perch (coastal)	MU (‘fillet) & LI	W				
PCB and dioxins and furans	Non-dioxin (PCB): PARGROUP = OC-CB	Biota	Herring & cod (open sea) Flounder & Perch (coastal)	MU (‘fillet) & LI	W				
PAH	PARAM = BAP	Biota	Molluscs & Crustaceans (M. edulis, M. baltica & Saduria entomon)	SB, TM	W				

² Denmark has a study reservation on secondary threshold EAC CB-118 24 µg/kg lw fish liver of muscle, the row is not included in the extraction table

Indicator	Parameters (PARAM) / Parameter groups (PARGROUP) (see also http://vocab.ices.dk/)	Primary matrix / GES	Species	Matrix	Basins	Secondary matrix / GES	Species	Matrix	Basins
PAH (fluoranthene)	PARAM = FLU					Sediment Biota	Molluscs & Crustaceans (M. edulis, M. baltica & Saduria entomon)	All SB	D W
PAH (secondary anthracene)	PARAM = ANT					Sediment		All	D
TBT and imposex	PARAM = TBTIN, TBSN+					Water			
TBT and imposex (included as test)	PARAM = TBTIN, TBSN+	Sediment		All	D				
TBT and imposex (included as test)	Imposex: PARAM = VDS, VDSI, INTS, INTSI, IMPF% Assisting parameters: PARAM = MBTIN, MBSN+, DBTIN, DBSN+, TBTIN, TBSN+, TPTIN, TPTSN+	Biota	Gastropods	All	D				

Autosubmission and resubmission facility

An autosubmission tool has been developed and are currently in use by Estonia, Sweden, Denmark, Poland and Germany. During summer 2017 the tool will be extended with auto submission capabilities. The tool enables Contracting Parties to automatically resubmit and submit data to the COMBINE database.

Revision of reporting format

A simplified version of the ICES Environmental Reporting format has been developed. The purpose of the simplified reporting format is to ease the reporting task for contracting parties by offering an optional reporting format that can be used in Excel or a similar spreadsheet.

Data output referencing through unique identifiers

ICES has implemented a system to allocate DOI (digital object identifiers) through DataCite. The Data product outputs and metadata will later be available online through the ICES publications library (<http://ices.dk/publications/library/Pages/default.aspx>) and will be referable through a DOI number.

Comparison of fields in COMBINE and EIONET

A comparison of COMBINE and EIONET dataformats was carried out in order to conclude if the data reported by some countries to EIONET can be used in the core indicator calculation if all relevant metadata is available. The overview below is based on the [ICES reporting format for environmental data](#). The fourth column indicates if the relevant field is covered by the draft EIONET reporting format.

It is important to emphasize that the EIONET data flow is focused on contaminants in seawater, sediment and biota as well as Oceanographic data. This implies that some of the ICES records are not relevant for comparison with the EIONET format. These records have therefore been omitted from this overview.

There are no **sampling and analytical methods**³ information in the EIONET data (except for Chlorophyll data). Furthermore information on analytical quality assurance and intercomparison, as well as uncertainty values are also missing in the EIONET reporting format. Therefore these record are not included in this overview.

Findings and conclusions

The column furthest to the right shows the conclusions for the HELCOM Expert Network on whether discrepancy in the metadata formats detected in the comparison are critical for indicator calculation purposes. It was overall concluded that for data to be useable in HELCOM assessments, it should be reported to COMBINE and not EIONET due to the detected discrepancies in the formats. The effort needed to convert and reformat the data should be taken into account when considering the added value of bringing in more data to the assessment. The COMBINE-EIONET data reporting comparisons seems to indicate that generally all data which is reported to EIONET is also reported to COMBINE. A general issue identified was that EIONET does not use so called 'methods records' which are considered as very important but not a critical question affecting all of the reported data in addition to the identified major fields.

³ records 20, 21 in the ICES format

Tables

RECID 00	File Information Record	8
RECID 90	Sampling Platform Record	9
RECID 91	Station / Sampling Event Record	10
RECID 03	Contaminant, Biological Effects AND LITTER Sample Record	12
RECID 04	Biota Specimen Data Record	14
RECID 10	Parameter Measurement Record	17

RECID 00 File Information Record

File Information Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
RECID	Record identifier	'00'	M	Not relevant	
RLABO	Reporting institute code	cf. RLABO	M	NO	This is a critical field - Traceability of data, to get back to the originator
CNTRY	Country code	cf. CNTRY	M	YES	
MYEAR	Monitoring year	YYYY	M	YES	

File Information Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
RFVER	Reporting format version number	‘3.2.5’	M	Not relevant	
DCFLG	Data centre flag - Reserved			Not relevant	

Sampling Location Records

RECID 90 Sampling Platform Record

Sampling Platform Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
RECID	Record identifier	‘90’	M	Not relevant	
SHIPC	Platform / Ship code	cf. SHIPC	M	(YES) Only nutrients data	Not critical - mainly relevant for small boats, maybe relevant for some parameters
CRUIS	Cruise identifier (series of sampling occasions) (must be unique for file)	Any character 0–9, A–Z etc.	M	NO	Not critical
Owner	Data owner	Any character 0–9, A–Z etc.	R	NO	Not critical

PRDAT	Public release date	Date YYYYMMDD	R	NO	Not critical
DCFLG	Data centre flag - Reserved			Not relevant	

RECID 91 Station / Sampling Event Record

Station Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
RECID	Record identifier	‘91’	M	Not relevant	
CRUIS	Cruise identifier (series of sampling occasions) (must be unique for file)	Any character 0–9, A–Z etc.	M	NO	Not critical
<i>STNNO</i>	Station identification /Sampling event ID	Any character 0–9, A–Z etc.	M	YES	
LATIT	Latitude (degrees/minutes/decimal minutes or as decimal degrees) Report as WGS84	–90 00.000 to +90 00.000 or -90.0000 to +90.0000 cf. Field Code Descriptions	M	YES	
LONGI	Longitude (degrees/minutes/decimal minutes or as decimal minutes) Report as WGS84	–180 00.000 to +180 00.000 or –180.0000 to +180.0000 cf. Field Code Descriptions	M	YES	
POSYS	Positioning system	cf. POSYS multiple options possible (separate multiple entries with “~” (ascii 126))	R	NO	Not critical
SDATE	Sampling date	YYYYMMDD	M	YES	

Station Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
STIME	Sampling time/start (UTC)	0000–2359 (hhmm)	R Fish disease	Not relevant	
ETIME	Sampling end time (UTC)	0000–2359 (hhmm) See field description for handling time when sampling continues to the next day.	R Fish disease	Not relevant	
WADEP	Water depth (sounding in meters)	0–9	rCW, rZB	YES	
STATN	Station name	Any character 0–9, A–Z etc. (max. 50)	R	YES	
MPROG	Monitoring programme	cf. MPROG multiple programmes possible (separate multiple entries with “~” (ascii 126))	M	(YES) Not mandatory and some MPROG’s are not part of the code list	Indication to CPs that they should include this information although it’s not mandatory
WLTP	Water/land types (river basin/eurotypes)	cf. WLTP	R	YES	
MSTAT	Type of monitoring station cf. Eurowaternet Technical report number 97	cf. MSTAT (replaces point source) (separate multiple entries with “~” (ascii 126))	R	(YES) All the MSTAT’s are not part of the code list	Yes this is a critical field - need to look at origin

Station Record – mandatory record –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
PURPM	Purpose of monitoring	cf. PURPM (separate multiple entries with “~” (ascii 126))	R	(YES) Purpose of monitoring have slightly different meanings in EIONET’s and ICES’s terminology	Not critical but helpful MIME approach filters for temporal
EDATE	Sampling end date	YYYYMMDD	mPassive samplers	NO	Don’t have passive sampling data
DCFLG	Data centre flag - Reserved			Not relevant	

RECID 03 Contaminant, Biological Effects AND LITTER Sample Record

Field Code <i>Foreign key in bold, primary key in italics</i>	Field name	Valid value	Mandatory/ Recommended/ Optional	Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
RECID	Record identifier	‘03’	M	Not relevant	

CRUIS	Cruise identifier (series of sampling occasions) (must be unique for file)	Any character 0–9, A–Z etc.	M	NO	
STNNO	Station identification /Sampling event ID	Any character 0–9, A–Z etc.	M	YES	
DTYPE	Data type	cf. DTYPE	M	YES	
SMPNO	Sample identification (for each species in haul, each sediment core, each sediment grab, each water bottle)	Any character 0–9, A–Z etc.	M	YES	
SMLNK	Sampling method link	1–999	R	NO	Critical for sediments and watersamples
ATIME	Actual time of sampling (UTC)	0000–2359 (hhmm)	r Tidal waters	YES	
NOAGG	Number of aggregated samples (hauls, sediment cores or grabs) taken to comprise sample	2–99		NO	Not critical
SPECI	Species of specimen	cf. ITIS , cf. WoRMS	mCF	YES	
RLIST	Reference code list used for species ID	cf. RLIST	mCF	NO	Not critical
FINFL	Factors potentially influencing guideline compliance and interpretation of data	cf. FINFL multiple options possible (separate multiple entries with “~” (ascii 126))	r Fish disease r CF	(NO) although these things can be reported in the [Remarks] field	Not critical
DCFLG	Data centre flag - Reserved			Not relevant	

RECID 04 Biota Specimen Data Record

Biota Specimen Data Record – mandatory record for CF –				Covered by the EIONET reporting format?	<i>To be completed by hazardous substance workshop</i> If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
RECID	Record identifier	‘04’	M	Not relevant	
CRUIS	Cruise identifier (series of sampling occasions) (must be unique for file)	Any character 0–9, A–Z etc.	M	NO	Not critical
STNNO	Station identification /Sampling event ID	Any character 0–9, A–Z etc.	M	YES	
SMPNO	Sample identification (for each species in haul, each sediment core, each sediment grab, each water bottle)	Any character 0–9, A–Z etc.	M	YES	
<i>SUBNO</i>	Sub-sample identification (each fish, egg, bird or aggregate pool of same species. New species – new sample record and SMPNO)	Any character 0–9, A–Z etc.	M	NO	Unclear if CPs are reporting pooled samples to EIONET, to be followed up by the CPs that report to inform on how the pooled samples are reported
NOINP	Number of individuals in sub-sample (i.e. 1 individual or number in pool)	01–99999	M	NO	See above
ORGSP	Origin of specimen	cf. ORGSP	r Imposex	NO	Not critical

Biota Specimen Data Record – mandatory record for CF –				Covered by the EIONET reporting format?	To be completed by hazardous substance workshop If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
SEXCO	Sex code	cf. SEXCO	R	NO	Not critical – but helpful in interpreting the assessment
STAGE	Stage of development	cf. STAGE	R	NO	Critical if <u>age</u> is not reported in any other way
CONES	Condition of specimen	cf. CONES	R	NO	Not critical
ASTSA	Animal state at time of sampling	cf. ASTSA	O	NO	Not critical
NODIS	Number of diseases looked for during a fish disease survey	0–99	r Fish disease	Not relevant	

Biota Specimen Data Record – mandatory record for CF –				Covered by the EIONET reporting format?	<i>To be completed by hazardous substance workshop</i> If previous column is “NO” is the major field required for HOLAS II data inclusion
Field Code <i>Foreign key in bold, primary key in italics</i>	Field Name	Valid Value	Mandatory/ Recommended/ Optional		
BULKID	Bulk identification (for individuals only)	If an individual (or parts thereof) has been analysed in one or more bulks, insert the SUBNO identification(s) of the bulk(s). Note that BULKID can only refer to a SUBNO within the same sample. See field descriptions for an example. (separate multiple entries with “~” (ascii 126))	r Fish disease r Biological effects data	NO	Not critical
DCFLG	Data centre flag - Reserved			Not relevant	

measurement records

RECID 10 Parameter Measurement Record

Field Code <i>Foreign key in bold, primary key in italics</i>	Field name	Value	Mandatory/ Recommended/ Optional	Covered by the EIONET reporting format?	<i>To be completed by hazardous substance workshop</i> If previous column is "NO" is the major field required for HOLAS II data inclusion
RECID	Record identifier	'10'	M	Not relevant	
CRUIS	Cruise identifier (series of sampling occasions) (must be unique for file)	Any character 0–9, A–Z etc.	M	NO	
STNNO	Station identification /Sampling event ID	Any character 0–9, A–Z etc.	M	YES	
SMPNO	Sample identification (for each species in haul, each sediment core, each sediment grab, each water bottle)	Any character 0–9, A–Z etc.	M	YES	
SUBNO	Sub-sample identification (each fish, egg, bird, aggregate, grab portion or core slice)	Any character 0–9, A–Z etc.	mCF mCS	NO	
MATRX	Matrix analysed	cf. MATRX	M	YES	
DEPHU	Upper depth/pressure (m) For pressure, see PRFLG in parameter record Upper depth of core slice for sediment.	0–9	R	YES	

Field Code <i>Foreign key in bold, primary key in italics</i>	Field name	Value	Mandatory/ Recommended/ Optional	Covered by the EIONET reporting format?	<i>To be completed by hazardous substance workshop</i> If previous column is "NO" is the major field required for HOLAS II data inclusion
DEPHL	Lower depth/pressure (m) See PRFLG in parameter record	0–9	R	(YES) For sediment	
PARAM	Parameter code	cf. PARAM	M	YES	
MUNIT	Measurement	cf. MUNIT	M	YES	
BASIS	Basis of determination	cf. BASIS	R for chemicals	YES	
AMLNK	Analytical methods link	1–999	Depends on parameter	NO	Very critical - used to conclude if samples are comparable
VFLAG	Validity flag	cf. VFLAG multiple flags possible (separate multiple entries with "~" (ascii 126))		NO	Not critical
QFLAG	Qualifier flag	cf. QFLAG multiple flags possible (separate multiple entries with "~" (ascii 126))		YES	
VALUE	Value	Any format	M	YES	

Field Code <i>Foreign key in bold, primary key in italics</i>	Field name	Value	Mandatory/ Recommended/ Optional	Covered by the EIONET reporting format?	<i>To be completed by hazardous substance workshop</i> If previous column is "NO" is the major field required for HOLAS II data inclusion
PERCR	Percentage recovery - to be applied (if thought necessary by data submitter) to the reported value (in VALUE field) at an assessment to give a better approximation of the real value	1–100	O	NO	Not critical
SIGND	Significant digits reported in VALUE	0–9		NO	Not critical
UNCRT	Uncertainty value	0–9	M (chemicals)	NO	Not critical but will lower the confidence of the assessment
METCU	Method of calculating uncertainty	cf. METCU	M	NO	See above
DETLI	Limit of detection value	0–9		YES	
LMQNT	Limit of quantification	0–9	R	YES	
PRFLG	Pressure flag	'Y' for Yes or blank		NO	Not critical
DCFLG	Data centre flag - Reserved			Not relevant	