SeaDataNet

FORMAT DOCUMENTATION

Recommended data transport format for biological data
in the framework of SeaDataNet
Deliverable D8.4b

Deliverable number		Short Title	
D8.4 (part B)		Transport format for biological data in SeaDataNet	
Long title			
Recommended data transport format for biological data in the framework of SeaDataNet			
Short description			
This document describes the proposed data transport format for biological data in the framework of SeaDataNet.			
Author		**Working group**	
K. Deneudt, Daphnis De Pooter		WP8	

History

Version	Authors	Date	Comments
1.0	Klaas Deneudt	28/11/2013	Creation
2.0	Daphnis De Pooter	26/11/2014	Update
2.1	Stéfane GOUZIEN	13/06/2019	Change information about [#] for P06 UUUU, and INDEXED-TEXT
2.2	Michèle.fichaut	01/06/2020	§5.2 change the answer to the question about unknown bottom depth Remove reference to Download manager

Contents

1. Introduction

One of the objectives of SeaDataNet II is to undertake actions to make SeaDataNet better fit for handling marine biological data sets and establishing interoperability with biology infrastructure developments. Based on an analysis of present biology data standards and initiatives, such as the OBIS, GBIF, TDWG and WoRMS standards, a recommended format for data transport of biological data is developed. This document describes the proposed format.

2. Objectives

The format should enable National Oceanographic Data Centers (NODC's) to make biological data accessible using SeaDataNet infrastructure and should make it possible for NODC's to use SeaDataNet to exchange biological data and to contribute to initiatives like (Eur)OBIS and GBIF.

This general purpose results in two specific requirements for the format:

- The format should be a general and higher level format without necessarily containing all specifics of each data type, but rather focusing on common information elements for marine biological data.

- At the same time the format needs to be sufficiently flexible/extendable to be applicable for at least part of the variety of biological data NODC's are managing.

- It should be possible to derive OBIS or Darwin Core compatible datasets from the format.

The format should be self-describing, in the sense that all information needed to interpret the data should be included in the file format or available through links to vocabularies or term lists that are part of the format.

While the primary target is data managers, it would be preferable to have a simple flat and human readable format that can as easily be created by scientists themselves starting from their original biological datasets which often reside in spreadsheets.

3. Format documentation for biology data

Accommodating data in the SeaDataNet infrastructure requires a detailed description of the data in the SDN metadata formats (CDI, CSR, EDMED, EDMERP, EDIOS) and storage of the data in one of the data transport formats (ODV, MedAtlas, NetCDF files).

For providing data access to biology data through the SDN portal the creation of CDI XML files storing the metadata and corresponding ODV files is the chosen approach. The CDI XML files for biology metadata are no different from the general CDI files. For the data itself, a specific variant of the ODV files for biology data is available and described below.

The format as described in this document is compatible with the ODV software (version 4.6.5). The data can be imported without error in ODV using the "import → SDN spreadsheet" tool. However ODV is not (yet) able to visualize the imported data or calculate statistics on them. ODV can however be used to aggregate the different SeaDataNet data files and export the data into a single spreadsheet. Note: This format can (at present) not be converted to netcdf CFPoint format.

3.1. CDI metadata

Since CDI XML files for biology metadata are not really different from other CDI files we can refer for the creation of these files to existing documentation on the CDI (ISO193139 model). http://www.seadatanet.org/Standards-Software/Metadata-formats/CDI

For certain types of biology data specific optional CDI fields might be required. For example:

TRACKS (Curves)   When describing line tracks by e.g. demersal fishing, this optional field can be used to describe the fishing track.

AREAS (Surfaces)   In case areas are surveyed instead of point stations, this optional field can be used to describe the multi-surface or polygon of each area entity.

The use of an EDMED REFERENCE in the CDI is highly recommended for biology data. EDMED is a dataset catalogue that can be used to create discovery metadata at a higher aggregate level. This field can be used to refer to the appropriate EDMED code referring to an item in the EDMED catalogue where a higher level description of the data collection can be found. Instructions for creating an entry for the EDMED catalogue can be found at http://www.seadatanet.org/Standards-Software/Metadata-formats/EDMED.

3.2. ODV biology variant

3.2.1. Mandatory fields

The general format of the SeaDataNet ODV import format is described at http://www.seadatanet.org/Standards-Software/Data-Transport-Formats.

The biology variant of the ODV file deviates from the general format in a number of ways. The format uses specific mandatory data parameter columns in addition to the mandatory elements already described in the Data Transport Format Manual. These biology mandatory fields are described below. Each of them has corresponding P01 concepts that should be included in the header.

It is important that **all mandatory fields** specified here are **present** (even if there is no data for some of the fields) and that the subject elements and column headers are **written exactly as specified** below.

A number of these mandatory fields are string variables. ODV will only correctly import the string variables, if a referral to the SDN:P06 conceptID UUUU in the <units></units> section of the header. Parameters for which the data types are integers or doubles should contain the appropriate unit in the column header, as well as in the semantic header.

ODV will only correctly import the string variables of optionnal columns, if the column headers of these fields **contain the unit suffix :INDEXED_TEXT** and a referral to the SDN:P06 conceptID UUUU in the <units></units> section of the header. For mandatory collumn INDEXED_TEXT suffix is optional for string variables. For mandatory collumn INDEXED_TEXT suffix is optional for string variables.

MinimumDepthOfObservation [m]/ MaximumDepthOfObservation [m]	Describes the minimum and maximum observation depths. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::MINWDIST*. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::MAXWDIST* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::ULAA*
SampleID	A unique identifier for a unit of material, or other medium of observation, gathered at a discrete point in time, which serves as a sample for a survey of biological occurrence (presence, quantification, absence, or derived value). The definition of a sample and assignment of IDs is controlled by the data originator. If applicable, details about how sample are defined and IDs assigned will be provided in metadata [original term from IOOSBiology]. *Refers to object: www.seadatanet.org/urnurl/SDN:P06::UUUU*
SamplingEffort [unit]*	The amount of effort expended during an Event. (It can be a linear distance (e.g. for a visual transect or net haul), a duration (e.g. for a entangling net) a volume (e.g. for a phytoplankton sample), a surface area (e.g. for a benthic core), etc.) [original term from DwC]. **For length sampled** *Refers to object: www.seadatanet.org/urnurl/SDN:P01::LENTRACK* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::ULAA* **For duration sampled** *Refers to: www.seadatanet.org/urnurl/SDN:P01::AZDRZZ01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UMIN* **For volume sampled:** *Refers to: www.seadatanet.org/urnurl/SDN:P01::VOLWBSMP*

Refers to unit: www.seadatanet.org/urnurl/SDN:P06::ULIT

For area sampled:
Refers to: www.seadatanet.org/urnurl/SDN:P01::AREABEDS
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UMSQ
* Replace "unit" with appropriate unit

ScientificName	The full name of lowest level taxon the Cataloged Item can be identified as a member of; includes genus, specific epithet, and subspecific epithet (zool.) or infraspecific rank abbreviation, and infraspecific epithet (bot.) Use name of suprageneric taxon (e.g., family name) if Cataloged Item cannot be identified to genus, species, or infraspecific taxon. Check spelling of names against World Register of Marine Species [original term from OBIS]. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::SCNAME01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*
ScientificNameID	An identifier for the nomenclatural (not taxonomic) details of a scientific name. [original term from DarwinCore]. Preferred format: urn:lsid:marinespecies.org:taxname:ID *Refers to object: www.seadatanet.org/urnurl/SDN:P01::SNANID01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*
Sex	The sex of a specimen or collected/observed individual(s). The domain should be a controlled set of terms (codes) based on community consensus. Proposed values: Male; Female; Hermaphrodite; Indeterminate (examined but could not be determined; Unknown (not examined); Transitional (between sexes; useful for sequential hermaphrodites); Both Male and Female [original term from OBIS]. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::ENTSEX01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*
LifeStage	Indicates the life stage present.. Can include multiple stages for a lot with multiple individuals. Preferably using the S11 controlled vocabulary: http://vocab.nerc.ac.uk/collection/S11/current/ [original term from OBIS]. *Refers to: www.seadatanet.org/urnurl/SDN:P01::LSTAGE01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*
ObservedIndividualCount	For the taxon under consideration, give the number of individuals that was found in the (sub)sample described by (sub)SampleID. The ObservedIndividualCount represents the number of individuals in the (sub)sample uniquely identified by the (sub)SampleID. If only presence data or aggregated density or biomass data is available, this field can be left empty (NULL). In case of providing absence information for a biological entity, use 0. [original term from OBIS]. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::OCOUNT01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*

As is generally the case for data variables in the ODV format, all the above data fields are paired with qualifying flags using the SeaDataNet vocabulary for qualifying flags (QV:SEADATANET). The Quality flag codes are available at *www.seadatanet.org/urnurl/SDN:L20*.

3.2.2. Conditional fields

In addition to the strictly mandatory fields a number of fields are considered conditional. This means they are mandatory in some predefined cases.

EventID

To be used **in case** an EventID would differ from the mandatory provided SampleID.

A unique identifier to group different samples of a unit of material which were gathered during a single deployment of an instrument. The definition of an event and assignment of IDs is controlled by the data originator. If applicable, details about how sample are defined and IDs assigned will be provided in metadata [original term from DarwinCore].
Refers to object: www.seadatanet.org/urnurl/SDN:P01::EVNTID01
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU

EventStartDateTime

To be used **in case** biological entities are observed or sampled during a recorded time period - not elsewhere given in metadata (e.g. CDI)

String corresponding to format 'YYYY-MM-DDThh:mm:ss.sss' or other valid ISO8601 string specifying the timestamp of the start of an activity. The detail included in the string represents the known precision. This field contains the startdatetime of the deployment of the instrument to which the EventID refers.
Refers to object: www.seadatanet.org/urnurl/SDN:P01::STRT8601
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU

EventEndDateTime

To be used **in case** biological entities are observed or sampled during a recorded time period that is not elsewhere described in metadata (e.g. CDI)

String corresponding to format 'YYYY-MM-DDThh:mm:ss.sss' or other valid ISO8601 string specifying the timestamp of the end of an activity. The detail included in the string represents the known precision. This field contains the Enddatetime of the deployment of the instrument to which the EventID refers.
Refers to object: www.seadatanet.org/urnurl/SDN:P01::ENDX8601
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU

EventStartLatitude / EventEndLatitude / EventStartLongitude / EventEndLongitude

To be used **in case** biological entities are observed or sampled during a recorded spatial track or line that is not elsewhere described in metadata.(e.g. CDI)

For samples/observations/record events better represented as line features rather than point features (e.g. extended trawls or transects) this indicates the starting/ending longitude/latitude location from which the specimen was collected or in which the sample/observation/record event occurred. This value should be expressed in decimal degrees (East & North = +; West & South = -). GPS-derived data must use the WGS 84 geodetic reference system (http://www.wgs84.com/) [original term from OBIS].
Refers to object: www.seadatanet.org/urnurl/SDN:P01::STRTXLAT
Refers to object: www.seadatanet.org/urnurl/SDN:P01::ENDXXLAT
Refers to object: www.seadatanet.org/urnurl/SDN:P01::STRTXLON
Refers to object: www.seadatanet.org/urnurl/SDN:P01::ENDXXLON
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU

SamplingProtocol

To be used **in case** results of different sampling or observation protocols are used within the same data file and information is not available elsewhere in the metadata (e.g. CDI)

The name of, reference to, or description of the method or protocol used during an Event. [original term from DWC]
Refers to object: www.seadatanet.org/urnurl/SDN:P01::SAMPPROT
Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU

SubsampleID	To be used **in case** subsampling occurred, subdividing the originally acquired sample.
	If a sample is further divided by the data originator for purposes of 0organization, handling, or analysis, and if the data originator wants to maintain the identity of the subsample, capture the ID of the subsample here. Explain in metadata if applicable [original term from IOOSBiology]. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::SSAMID01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*
SubSamplingCoefficient	To be used **in case** subsampling occurred, subdividing the originally acquired sample.
	A decimal coefficient lower or equal to 1 that indicates the proportion of the Sample that is represented in the Subsample. This value allows rescaling the counted (and measured) individuals to the total sampling effort applied at Sample level. If no proportional subsampling occurred this value should equal 1. *Refers to object: www.seadatanet.org/urnurl/SDN:P01::SSAMC01* *Refers to unit: www.seadatanet.org/urnurl/SDN:P06::UUUU*

3.2.3. Optional fields

On top of the mandatory and conditional fields, the biology variant allows the use of additional optional fields. These fields can be selected depending on the data type under consideration and must refer to available concepts in the SeaDataNet P01 and P06 vocabularies (in the header of the file, see 3.2.3).

Examples of these additional fields are:

- DensityPerUnitEffort [#/m²]
www.seadatanet.org/urnurl/SDN:P01::SDBIOL02
(definition: Abundance of unspecified biological entity per unit area of the bed)
www.seadatanet.org/urnurl/SDN:P06:PMSQ
(definition: per square metre)

- DensityPerUnitEffort [#/l]
www.seadatanet.org/urnurl/SDN:P01::SDBIOL01
(definition: Abundance of unspecified biological entity per volume of the water body)
www.seadatanet.org/urnurl/SDN:P06:UCPL
(definition: number per litre)

- WWBiomassPerUnitEffort [kg/m²]
www.seadatanet.org/urnurl/SDN:P01::SDBIOL05
(definition: Wet weight biomass of unspecified biological entity per unit area of the bed)
www.seadatanet.org/urnurl/SDN:P06:KMP2
(definition: kilograms per square metre)

- ObservedIndividualLength [cm]
http://www.seadatanet.org/urnurl/SDN:P01::OBSINDLX
(definition: Length of unspecified biological entity)
http://www.seadatanet.org/urnurl/SDN:P06::ULCM
(definition: centimetres)

You can assign names to the column headers of optional fields as preferred, however there are some guidelines which need to be kept in mind. It is not advisable to use spaces, brackets or punctuation in the column header names. In contrast, underscores are allowed. Note that in any given file the names used as column header must be unique.

Also string type parameters are possible as optional fields, if they cannot be avoided by using a coding convention (e.g. 0=no / 1= yes). The column header of optional string type parameters should contain

the suffix ":INDEXED_TEXT" (without a space between the parameter name and the suffix). This suffix will allow Ocean Data Viewer to correctly interpret this field as a string.

In order to illustrate the possibilities, the following concepts have been created to be used for specific types of biology data (abundance and biomass data, data for specimen characteristics, ecotoxicology data, data on habitat characteristics, …). A similar approach can be used to extend to other biological data types.

3.2.3.1.   Concepts for abundance and biomass

- Wet weight biomass (in assayed sample) of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::OWETBM01

- Abundance of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL01

- Abundance of unspecified biological entity per unit area of the bed
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL02

- Ash-free dry weight biomass of unspecified biological entity per unit area of the bed
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL03

- Wet weight biomass of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL04

- Wet weight biomass of unspecified biological entity per unit area of the bed
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL05

- Abundance category of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL06

- Ash-free dry weight biomass of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL07

- Dry weight biomass of unspecified biological entity per unit area of the bed
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL08

- Dry weight biomass of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL09

- Coverage (in assayed sample) of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL10

- Coverage (HELCOM COMBINE) of unspecified biological entity on the bed
http://www.seadatanet.org/urnurl/SDN:P01::HELCCVRG

- Count (breeding pairs) of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::SOVO0001

- Count (January) {midwinter count} of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::SOVO0004

- Maximum count (seasonal) of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::SOVO0003

- Mean count (seasonal) of unspecified biological entity

http://www.seadatanet.org/urnurl/SDN:P01::SOVO0002

- Calculated carbon biomass of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL12

- Calculated volume of unspecified biological entity per unit volume of the water body
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL13

3.2.3.2. Concepts for specimen characteristics

- Specimen age of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::AGEBENTX

- Categorisation of unspecified biological entity as living or non-living
http://www.seadatanet.org/urnurl/SDN:P01::SDBIOL11

- Length of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::OBSINDLX

- Maximum length of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::OBSMAXLX

- Minimum length of unspecified biological entity
http://www.seadatanet.org/urnurl/SDN:P01::OBSMINLX

3.2.3.3. Concepts for ecotoxicology

- Concentration of lead {Pb} per unit wet weight of biota {unspecified biological entity}
http://www.seadatanet.org/urnurl/SDN:P01::PBBIOTUK

3.2.3.4. Concepts for habitat characteristics

- Sediment type category
http://www.seadatanet.org/urnurl/SDN:P01::SEDTYCAT
The data type of this parameter is a string. The column header should therefore contain the ":INDEXED_TEXT" suffix: SedimentTypeCategory:INDEXED_TEXT

Proportion of particles (sand size-fraction) in sediment
- http://www.seadatanet.org/urnurl/SDN:P01::SANDZZZZ

Proportion of particles (silt size-fraction) in sediment
- http://www.seadatanet.org/urnurl/SDN:P01::SILTZZZZ

Proportion of particles (clay size-fraction) in sediment
- http://www.seadatanet.org/urnurl/SDN:P01::CLAYZZZZ

If your dataset contains a biological parameter which is not yet represented in this list please contact Daphnis.Depooter@vliz.be with your request. A suitable concept will be created in correspondence with the vocabulary management team at BODC.

3.2.4. Semantic header

The semantic header of the ODV biology variant follows the general ODV semantic header rules as outlined in the Data Transport Formats manual at http://www.seadatanet.org/Standards-Software/Data-Transport-Formats.

The semantic header is needed as a mapping that provides a reference for every data column heading label to standardized vocabulary concepts.

Notice that in the newest version an <instrument> element is allowed in addition to the <object>, <subject>, <units> element. **The use of this <instrument> element referring to the L22 vocabulary is recommended for biology data.**

All available concept codes can be found using the user interfaces at:
- Parameter: http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P01
- Units: http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=P06
- Instruments : http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L22

4. Examples of biology data

For purposes of demonstration, a template has been developed in a spreadsheet file (**SDN_Biol2.0 template**). This template contains a tab with all **mandatory fields** in the correct format, and lists all **optional fields** which have been requested in the so far in the framework of providing biology data using the ODV biology format. Two fictitious example datasets have been worked out in the same spreadsheet file.

5. FAQ

5.1. New conceptID's

Q) I cannot find an appropriate conceptID for a parameter (or instrument) used in my dataset.

A) Contact Daphnis.Depooter@vliz.be with a detailed description of the parameter you need. In case of an instrument we'll also need a paper describing the parameter or a reference to the manufacturer.
We'll then ask our contact at BODC to add it for you and we'll keep track of all conceptid's which have been requested in the framework of the biological format for SeaDataNet, so they can be reused by other partners.

Q) I would like to reference a specific sampling methodology rather than an instrument. Is this possible?

A) Yes, a parameter was added just for this. It's called SamplingProtocol (conceptID SAMPPROT). It's preferred if you add in this field a detailed description of the protocol or add a reference to a paper. If you wish to add a URL to a publication, it is preferred that you add a stable URL (like a DOI, or a link to a repository). If this in not available you can contact Daphnis.Depooter@vliz.be and we'll provide a stable URL to the publication in the VLIZ publication catalogue.
The sampling protocol concept does not list gears but refers to a protocol.

5.2. Questions about specific conceptID's

Q) What is the **difference** between **Local_CDI_ID, EventID, SampleID and subsampleID**?

A) **Local_CDI_ID**: In theory this is a unique code for each sampling event. Each deployment of a grab/trawling device should be describe using a different CDI. It can also group a time series of samples taken in (direct) succession at a given location. The SeaDataNet Biology template also allows a single CDI to describe the long term monitoring results at a single station, provided that the methodology did not differ. The local_CDI_ID's may even contain data from different stations, although this is not recommended. In this case it is important to have a good look at the conditional fields.

EventID: An EventID is an ID that uniquely defines the action in the field. It is used to group data gathered during the same Event. For example an event is a single deployment of a grab/trawling device or an observation action. When the same device is deployed two times at single station, this results in different EventIDs. EventID is consided a conditional field, that should be used whenever it is applicable.

SampleID: One Event can result in different samples, which should each be described by different SampleID's. For example one vertical net deployment can gather different samples from different depth intervals.

SubsampleID: Sometimes (trawling data, sediment grabs, …) only a representative part of a sample which was collected is analysed. Based on an estimate of the portion of this sample,

interpolations are made to calculate the values for the entire sample. The SubSampleID allows grouping of this raw data. It's always to be used together with the SubSamplingCoefficient. If in your dataset the entire sample was analysed, you should use the same id for SampleID and SubSampleID and add 1 as SubSamplingCoefficient. SubsampleID is consided a conditional field, that should be used whenever it is applicable.

Q) My Local_CDI_IDs only contain data from a single sample taken from a single sampling event and my dataset contains no subsamples. **Can I leave** EventID, SampleID and SubSampleID and SubSamplingCoefficient **empty**?

A) EventID, SubsampleID, SubSamplingCoefficient are considered conditional fields, that should be used whenever they are applicable. Check 3.2.2 of the guidelines for more information on these conditions.
If some records in your dataset do not contain any subsamples, you can use the same code for SubsampleID as you used as a SampleID. The SubSampling coefficient should in this case be "1".

Q) I already provided the Start/EndDateTime and coordinates to the CDI record. Is it necessary to add the same information in the data file as EventStart/StopDateTime and EventStart / StopLongitude / Latitude ?

A) The ODV biology template also allows a single CDI to describe the long term monitoring results at a single station, provided that the methodology did not differ. In this case there is need for detailed information on the sampling event in the data file itself. Use the conditional fields described under 3.2.2. to do this.

Q) The **scientific name** in my dataset **differs** from the (**accepted) name** proposed by the **WoRMS** database. What should I do?

A) You should **<u>ALWAYS keep the original scientific name</u>** as it occurs in your database, even if the matching tool indicates that this name is not accepted, or if it contains indicators of uncertainty (such as "aff" or "var"). You should also provide the AphiaID of the direct match as the ScientificNameID even if this one is currently unaccepted according to WoRMS. The currently accepted name can always be retrieved using the (unaccepted) aphiaid through the WoRMS database. Keeping the original scientific name allows to correct possible errors occurring during the matching.

If in your database the ScientificName also contains the lifestage, gender, sizeclass, (specimen length or …) you will need to split these up into the appropriate fields.

Q) What is the **TYPE code** in the third column for the biological data?
A) The type code is *.

Q) In my dataset the **Bottom depth** is unknown or not applicable. What should I fill out?

A) If the Bottom depth is unknown you have to leave the field empty

Q) What does field **Bot. depth** indicate when samples are taken during a transect with different stops?

A) Bot. depth should be the bathymetrical dept. When you are covering a transect and record data along different stops in the transect, you have 2 options.

1) If you recorded the coordinates for each stop, then you can also add the bathymetrical depth at each stop. If you did not record the coordinates of each "stop"

2) If you do not have coordinates for each stop, you should use the bathymetrical depth at the starting position of transect for which you have the coordinates.

Q) Some samples in my dataset do not contain any reference towards the **depth** where the sample is taken. What should I do?

A) If you do not know the depth of sampling or if this is not applicable you can leave this field empty and assign the QF 9.

Q) Can I use abbreviations for **lifestage** or is there a **controled vocabulary**?

A) For clarity reasons it is preferred that the full text lifestage is given.
A controlled vocabulary for Lifestages called S11 is maintained at BODC, http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=S11. This one is used preferentially. Alternative vocabularies include the one maintained by ICES for its eggs and larvae dataset which might be useful (called "stage" at: http://vocab.ices.dk/).

Q) Is there a controlled vocabulary for gender ?

A) Yes, standardized values are proposed within the OBIS data schema, which are preferably written in full:
Male; Female; Hermaphrodite; Indeterminate (examined but could not be determined; Unknown (not examined); Transitional (between sexes; useful for sequential hermaphrodites); Both Male and Female

Q) For the mandatory **field Cruise**. My dataset did not result from a cruise. What should be used here?
A) You can use an identifier for the survey during which the data was gathered. Alternatively this can be the (shortened) name of the dataset.

5.3. General questions about ODV bio variant

Q) What should I do when my **sample did not contain any organisms**?

A) We recommend that you add a record containing the - for the dataset - highest relevant taxonomic level in the field ScientificName (http://www.marinespecies.org/aphia.php?p=browser&id=2&expand=false#ct), and add 0 as the ObservedIndividualcount and QF 1 for good value. E.g. Pisces, Animalia,…
Q) What should I do if the taxon was recorded, but **I don't have an ObservedIndividualCount**?

A) If you don't have an original count but you know the taxon was present, you can assign a NULL value (=leave blank) in combination with a QF of 1 or 2 for (probably) good value.

If the original count was lost, but you have a calculated abundance, you can still assign the NULL value to ObservedIndividualCount if the taxon was present, in combination with a parameter "DensityPerUnitEffort".

Do not assign the value 0 unless you want to provide absence information. A value of 0 (Zero) in the observerindividualcount would mean that the sample was examined specifically for this taxon, but it was not encountered.

Q) What should I do if my dataset does **not contain any information for a required field** of the template?

A) The general rule for ODV is that you can leave empty mandatory fields setting the QC flag that accompanies each column to QC value 9. QC value 9 means "missing value".

There are however a few exceptions:

For **Gender** and **Lifestage** which use a controlled vocabulary is used, there are an "unknown" and a "Not applicable" value available.

ObservedIndividualCount you can leave blank (NULL value) in combination with the QF 1 or 2 to indicate presence.

Q) My dataset contains both length measurements of fishes and a CatchPerUnitEffort (CPUE). Should these be split up into different data files?

A) The data can be split over different files, but there is another approach that we would recommend. There is a conceptid OBSINDLX to indicate the length of the specimen. You can list each specimen and its lengths separately. In principle, the abundance over all lengths can easily be calculated by summing the observedindividualcounts, divided by the samplingeffort.

If not for all specimens the length was recorded (e.g. because lengths were only measured until a normal distribution was found), you could add an additional record with the observedindividualcount of the specimens for which no length was recorded.

If the latter is unknown, you can add an additional record for the catch per unit effort. In this record the ObservedIndividualCount field should have the NULL value.

Q) For benthos sampling, there are situations where **only a part of an organism was encountered**. We could record this as for example 0.1 specimen, resulting in a decimal value for ObservedIndividualCount. Is that allowed?

A) If you encountered a piece of a specimen, you can leave blank (NULL value) to indicate presence.

Q) In the new Biological data format, is it necessary to apply for **new parameter conceptid for each taxon** recorded in my dataset?

A) No, in the unmodified ODV format it would have been necessary to use terms like you have given to label each data column. However, one of the main ideas of the ODV biology variant is

that the species name can be separated out as a parameter. The ODV biology format therefore contains a conceptid ScientificName and ScientificNameId, which can be used to store the original scientific name and the aphia LSID. So where before you would have needed a conceptid called "*Abundance of Abra (ITIS: 81301: WoRMS 138474) per unit area of the bed*" you can now use "*Abundance of unspecified biological entity per unit area of the bed*", and specify the taxon related information in the designated fields.

Q) For ecotoxicological datasets I have results of different types of organochlorines and heavy metals for different taxa in different genders and lifestages. How should this information be stored according to the ODV biology format?

A) The inefficient formatting these types of data is one of the concerns specifically addressed during the development of the biology ODV format. The SeaDataNet Biology template allows the storage of taxa, sex and lifestages separate from the contaminant components.

Contaminants fit into this format - what is needed are P01 codes along the lines of 'Concentration of lead {Pb} per unit wet weight of unspecified biota' to head the columns. A separate P01 for the matrix could be created to specify the body tissue in which the measurements took place.

Q) I try to import the ODV biology data file into ODV, but the data for the string values is missing.

A) Check that you downloaded the most recent version of the ODV. It should contain a fix for this problem.

Otherwise check in the downloaded files that the text parameters headers look exactly (mind spaces and capitalisation) like:

MinimumObservationDepth [m]
MaximumObservationDepth [m]
SampleID
ScientificName
ScientificNameID
Sex
LifeStage
ObservedIndividualCount
EventID
EventStartDateTime
EventEndDateTime
EventStartLongitude
EventEndLongitude
EventStartLatitude
EventEndLatitude
SubsampleID
SubSamplingCoefficient
Samplingprotocol:INDEXED_TEXT

The obsolet suffixes [#] and [none] to indicate that these parameters don't have a unit are supported, but others aren't. If EventID or SubsampleID are text field, don't forget :INDEXED_TEXT suffix after header (EventID:INDEXED_TEXT and SubsampleID:INDEXED_TEXT)

Q) How can I create an ODV file where the datetime field represents the known precision (e.g. only defining a year and a month)?

A) A fix has been approved to deal with this problem and will be implemented early 2015. The fix fill allow you change the description of the time_ISO8601. If you only have a year and Month then put time_iso8601 [YYYY-MM], if you only have year, month and day then time_iso8601 [YYYY-MM-DD] etc... . The string into [] represents the ISO description of date-time.

Q) What is the conceptID for SamplingEffort?

A) There are at least 4 different Concept id's which can be used as a sampling effort:

VOLWBSMP: a volume of the water body (e.g. when a bottle was used)

AREABEDS: A surface area (e.g. when a grab was used)

LENTRACK: transect length or the distance a net was towed through the water. This can be used in combination with **ZDRZZ01** to specify the duration of the towing

ZDRZZ01: a sampling duration, for instance to indicate the duration that standing net was deployed.

Others will become available if needed. Contact Daphnis.Depooter@vliz.be, if you need another sampling effort type.

Q) Would it not make more sense to split the sampling size (or other variables) in separate columns for the value and the unit?

A) This is inherent to the ODV format, units are designated per column (variable). There are only 2 options to deal with this issue. First option is standardizing the units. A second option is to put data with different sampling efforts in different data files. For a better understanding please read paragraph 2.1 of the SeaDataNet ODV manual. The ODV software offers the functionality to compile the different data columns and do a conversion of units while integrating the different data files.

Q) My dataset results from sampling with a Van Veen grab with a surface area of 0.20 m². How do I describe this in the new ODV biology format?

A) In the ODV biology format, this information is described in a separate fields:

1) ODV biology contains a mandatory field called **SamplingEffort**, which should be used to store the surface area sampled, volume sampled, track length of towing, etc…
For the example you should refer to the P01 conceptID AREABEDS for area, the UMSQ unit from the P06 and add the 0.20 value to all records.

2) You should select the appropriate conceptID for your instrument from the **SeaVoX Device Catalogue** http://seadatanet.maris2.nl/v_bodc_vocab_v2/search.asp?lib=L22.
The L22 conceptID should be included in the semantic header linked to the parameter ObservedIndividualCount, and/or other parameters indicating the abundance in your sample.

3) If a user could have any doubts on the methodolgy or instruments from which the data resulted, additional information should be added in the mandatory field **SamplingProtocol**. This field can contain a textual description of the methodology or a reference to a paper with a methodology. If no stable URL is available to refer to a paper, you can send the paper daphnis.depooter@vliz.be who will provide you one.