





European Marine Observation and Data Network

Biological data management

Flanders Marine Institute





Overview of biological data format



- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).
- Darwin Core (DwC) standard <u>terms</u>.
- Minimum of fields required per table.
- Date/time & Lat/lon.
- EventID and OccurrenceID.
- Taxonomic information: <u>LSID</u>
- Other parameters: <u>BODC-NERC vocabulary</u>.



()1

Data schema / structure

- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).

Data is structured in 1 to 3 tables related to each other. Number of tables depends on the nature of the data. This structure allows to store not only occurrences but also sampling information and additional biological and/or abiotic measurements. The general content of each table:

Event table	Occurrence table	Measurements or Facts (eMoF)
Sample and/or observation information (time, location, depth, event hierarchy)	Occurrence details (taxonomy, identification, organismID)	Sampling protocol (equipment, methods) Sampling effort (length, duration, volume) Environment/habitat variables (physical, chemical, sediment) Biological variables (Abundance, biomass, lifestage, sex)



01

Data schema / structure

- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).





()1

Data schema / structure

- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).

Occurrence table eventID occurrenceID The three tables are related via the Occurrence details eventID and the occurrenceID. (taxonomy, identification) Event table The eMoF Extension is used in combination with eventID the Event Core and the Occurrence Extension to capture both abiotic measurements and biotic Sample or Observation measurements. The occurrenceID is used to link (time, location, depth, event **Extended** hierarchy...) biotic measurements in the eMoF Extension with Measurement or Facts the the Occurrence Extension and the eventID eventID links the eMoE to the Event Core. occurrenceID Sampling protocol Sampling effort Environment/habitat variables **Biological variables**



Data schema / structure

- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).

When to use Event Core

- When the dataset contains abiotic measurements, or other biological measurements which are related to an entire sample (not a single specimen).
- When specific details are known about how a biological sample was taken and processed. These details can be expressed using the eMoF and the newly developed Q01 vocabulary.
- Event Core should be used in combination with the Occurrence Extension and the eMoF.

When to use Occurrence Core

- No information on how the data was sampled or samples were processed.
- No abiotic measurements are taken or provided.
- Biological measurements are made on individual specimens (each specimen is a single occurrence record) This is often the case for museum collections, citations of occurrences from literature, individual sightings. Datasets formatted in Occurrence Core can use the eMoF Extension for biotic measurements or facts.



Field nomenclature

- Darwin Core (DwC) standard <u>terms</u>.
- Minimum of fields required per table.

The field names of each of the 3 tables have to follow the Darwin Core terminology. The DwC terms that are most relevant to EMODnet Biology format are the following (those in **bold** are mandatory):

Event table

eventID, parentEventID, eventDate, habitat, minimumDepthInMeters, maximumDepthInMeters, decimalLatitude, decimalLongitude, coordinateUncertaintyInMeters, footprintWKT, modified

Occurrence table

eventID, occurrenceID, scientificName, scientificNameAuthorship, scientificNameID, kingdom, taxonRank, identificationQualifier, occurrenceStatus, basisOfRecord, modified

Extended MeasurementorFact table

measurementID, **eventID**, occurrenceID, measurementType, measurementTypeID, measurementValue, measurementValueID, measurementUnit, measurementUnitID, measurementAccuracy, measurementRemarks



Content - Controlled vocabulary and standards

- Date/time & Lat/lon.
- EventID and OccurrenceID.
- Taxonomic information: LSID
- Other parameters: <u>BODC-NERC vocabulary</u>.

Besides the field names, the content or the data itself has to follow certain standards. For example, the date-related fields have to be ISO 8601 compliant, the latitude and longitude have to be in decimal degrees and referenced to the WGS8484 datum.

An overview of the required format for the content of the different fields is available <u>here</u>.

In this document we describe how to format some the most relevant fields:

- EventID and OccurrenceID
- Taxonomic information: LSID
- Other parameters (eMoF table): BODC-NERC controlled vocabulary.





eventID

An identifier for the set of information associated with a (sampling) event.

с	D	E	F	G	н	i l	J	к
datasetName	eventID	arentEventID	eventDate	locationID	minimumDepthInMeters	s maximumDepthInMeters	decimalLatitude	decimalLongitude
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M1	EIOFUN1	30/05/2009	WM1200	1200	1200	38.3881	1.8165
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M2	EIOFUN1	30/05/2009	WM1200	1200	1200	38.3915	1.8125
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M3	EIOFUN1	01/06/2009	WM2000	2000	2000	38.038	1.9027
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M4	EIOFUN1	01/06/2009	WM2000	2000	2000	38.0482	1.9357
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M6	EIOFUN1	04/06/2009	WM3000	2800	2800	38.413	5.4199
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M8	EIOFUN1	07/06/2009	CM1200	1200	1200	38.2453	16.6663
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M9	EIOFUN1	07/06/2009	CM1200	1200	1200	38.2453	16.6663
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M10	EIOFUN1	08/06/2009	CM1200	1200	1200	38.2453	16.6663
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M13	EIOFUN1	15/06/2009	EM1200	1200	1200	34.5438	25.775
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M14	EIOFUN1	16/06/2009	EM1200	1200	1200	34.5438	25.775
Megafaunal data from the 2008 BIOFUN trans-Mediterranean deep-sea cruise	BF1M15	EIOFUN1	16/06/2009	EM1200	1200	1200	34.5438	25.775





occurrencelD

This field is a unique identifier for each record within the dataset or collection. But it has to be unique within the whole database. To ensure the uniqueness, it is usually made of a string concatenation which has three parts:

- Institute code / cruise code: e.g. BIOFUN1 (cruise), IPMA (institution)
- A unique identifier for the dataset/collection/sample:
 - e.g. Historical_Zoobenthos_Constanta_1997_1999 (dataset)
 - e.g. BF1M1 (sample)
- A number: 1, 2, 3, etc. -> to make each record unique within the dataset

	A	В	с	D	E	F	
1	id	basisOfRecord	ord occurrenceID		scientificNameID	scientificName	
2	BF1M1	MaterialSample	BIOFUN1_BF1M1_1	BF1M1	urn:lsid:marinespecies.org:taxname:126684	Alepocephalus rostratus	
3	BF1M1	MaterialSample	BIOFUN1_BF1M1_2	BF1M1	urn:lsid:marinespecies.org:taxname:299942	Bathypterois mediterraneus	
4	BF1M1	MaterialSample	BIOFUN1_BF1M1_3	BF1M1	urn:lsid:marinespecies.org:taxname:280313	Coelorinchus mediterraneus	
5	BF1M1	MaterialSample	BIOFUN1_BF1M1_4	BF1M1	urn:lsid:marinespecies.org:taxname:105812	Galeus melastomus	
6	BF1M1	MaterialSample	BIOFUN1_BF1M1_5	BF1M1	urn:lsid:marinespecies.org:taxname:126495	Lepidion lepidion	
7	BF1M1	MaterialSample	BIOFUN1_BF1M1_6	BF1M1	urn:lsid:marinespecies.org:taxname:126497	Mora moro	





aphialD: taxonomic information

All the occurrences are given a unique scientificNameID. This is done by matching the Scientific Names of your occurrence table with the <u>World Register of Marine Species</u>, using the <u>taxon match tool</u>. Information on how to use the taxon match tool <u>here</u>.

After matching, the tool will return you a file with the AphiaIDs, LSIDs, valid names, authorities, classification and any other output you have selected. The WoRMS **LSID** is used for DwC field **scientificNameID**.

C D		E	F		
occurrenceID	eventID	scientificNameID	scientificName		
BIOFUN1_BF1M1_1	BF1M1	urn:lsid:marinespecies.org:taxname:126684	Alepocephalus rostratus		
BIOFUN1_BF1M1_2	BF1M1	urn:lsid:marinespecies.org:taxname:299942	Bathypterois mediterraneus		
BIOFUN1_BF1M1_3	BF1M1	urn:lsid:marinespecies.org:taxname:280313	Coelorinchus mediterraneus		
BIOFUN1_BF1M1_4	BF1M1	urn:lsid:marinespecies.org:taxname:105812	Galeus melastomus		
BIOFUN1_BF1M1_5	BF1M1	urn:lsid:marinespecies.org:taxname:126495	Lepidion lepidion		
BIOFUN1_BF1M1_6	BF1M1	urn:lsid:marinespecies.org:taxname:126497	Mora moro		

Important: After the match, don't provide the "accepted" scientificName and scientificNameID but the matches of the original taxa from the data.





• The eMoF extension is used to store:

- information related to sampling method and sampling effort (via eventID).
- measurements linked to a biological occurrence (via occurrenceID).
- environmental measurements (via eventID).
- The MoF terms: measurementType, measurementValue and measurementUnit are completely unconstrained and can be populated with free text annotation.
 - While free text offers the advantage of capturing complex and as yet unclassified information, the inevitable semantic heterogeneity (e.g. of spelling or wording) becomes a major challenge for effective data integration and analysis.





- Three fields to standardise the measurement types, values and units: measurementTypeID, measurementValueID and measurementUnitID.
- These terms are populated using controlled vocabularies from the NERC Vocabulary Server, developed by the British Oceanographic Data Centre (BODC)
 - <u>https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/</u>

MeasurementType	MeasurementTypeID
(free text)	(controlled vocabulary)
Body length	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length (mm)	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
length_in_mm	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length of specimen	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX

Example of parameter standardization using controlled vocabulary.





MeasurementTypeID

 BODC Parameter Usage Vocabulary (P01): <u>http://vocab.nerc.ac.uk/collection/P01/current</u>
OBIS sampling instruments and methods attributes (Q01): <u>http://vocab.nerc.ac.uk/collection/Q01/current/</u>

MeasurementValueID

•Sampling instruments and sensors (SeaVoX Device Catalogue): <u>http://vocab.nerc.ac.uk/collection/L22/current</u>

•Sampling instrument categories (SeaDataNet device categories): <u>http://vocab.nerc.ac.uk/collection/L05/current</u>

Vessels (ICES Platform Codes):

http://vocab.nerc.ac.uk/collection/C17/current

•Lifestage: <u>http://vocab.nerc.ac.uk/collection/S11/current/</u> •DOIs of papers or manuals on the sampling protocol used, published e.g. on IOC's <u>Ocean Best Practices repository</u>, for example: <u>http://hdl.handle.net/11329/304</u>

MeasurementUnitID

·Units: http://vocab.nerc.ac.uk/collection/P06/current







id	Occurrence ID	Measurement Type	Measurement TypeID	Measurement Value	Measurement Unit	Measurement UnitID
BF1M1	BIOFUN1_BF1M1_1	Observed Individual Count	http://vocab.nerc.ac.uk/collection/P01/current/OCOUNT01	26	Individuals	http://vocab.nerc.ac.uk/collection/P06/current/UUUU
BF1M1	BIOFUN1_BF1M1_2	Observed Individual Count	http://vocab.nerc.ac.uk/collection/P01/current/OCOUNT01	3	Individuals	http://vocab.nerc.ac.uk/collection/P06/current/UUUU
BF1M1	BIOFUN1_BF1M1_1	Abundance	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL02	0.000329114	N/km2	http://vocab.nerc.ac.uk/collection/P06/current/NPKM
BF1M1	BIOFUN1_BF1M1_2	Abundance	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL02	3.80E-05	N/km2	http://vocab.nerc.ac.uk/collection/P06/current/NPKM
BF1M1	BIOFUN1_BF1M1_2	Wet Weight Biomass	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL05	0.000721519	kg/km2	

Example of measurements linked to a biological occurrence (via occurrenceID)

id	occurrenceID	measurementType	measurementTypeID	measurementValue	measurementAccuracy	measurementUnit
BF1M15		Trawling speed	http://vocab.nerc.ac.uk/collection/P01/current/APSAZZ01	2.7	0.1	knots
BF1M17		Trawling speed	http://vocab.nerc.ac.uk/collection/P01/current/APSAZZ01	2.7	0.1	knots
BF1A01		Gear	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100002	Agassiz dredge		
BF1A02		Gear	http://vocab.nerc.ac.uk/collection/001/current/00100002	Agassiz dredge		

Example of information related to sampling method and sampling effort (via eventID)



Summary and dataset example





Overview of biological data format



- 1-3 tables (<u>OBIS-ENV data format</u>):
 - Event core table.
 - Occurrence table.
 - Extended Measurements or Facts (eMoF).
- Darwin Core (DwC) standard <u>terms</u>.
- Minimum of fields required per table.
- Date/time & Lat/lon.
- EventID and OccurrenceID.
- Taxonomic information: <u>LSID</u>
- Other parameters: <u>BODC-NERC vocabulary</u>.





- Event table: sample and/or observation information (time, location, depth, event hierarchy)
- Occurrence table: Occurrence details (taxonomy, identification, organismID...)
- Extended MeasurementsOrFacts table: information related to sampling method and sampling effort; measurements linked to a biological occurrence; environmental measurements.
- The field names follow the DarwinCore terminology. Each of the 3 tables has a minimum number of fields required.
- The way the data is populated follows different standards, depending on the field. Some fields are free text, some others have controlled vocabulary:
 - Date/time -> ISO
 - Taxonomy -> WoRMS LSID from taxon match
 - eMoF -> NERC vocabulary
 - o etc.





Dataset example

Full dataset can be found at:

https://docs.google.com/spreadsheets/d/1b-RHC68IsqqggHK8ybgdnG3E_5CL0mum6v_Rn5hjkjU/edit?usp=sharing_

Published version at: http://ipt.vliz.be/upload/resource?r=biofun2008



Dataset example

D	D E		E F		GH I J		К	
eventID	parentEventID	eventDate		decimalLatitude	decimalLongitude			
BF1M1	BIOFUN1	30/05/2009		38.3881	1.8165			
BF11/12	BIOFUN1	30/05/2009		38.3915	1.8125			
BF11/13	BIOFUN1	01/06/2009		38.038	1.9027			

C I		E	F	
occurrenceID	ever tID	scientificNameID	scientificName	
BIOFUN1_BF1M1_1	BF1M1	urn:lsid:marinespecies.org:taxname:126684	Alepocephalus rostratus	
BIOFUN1_EF1M1_2	BF1M1	urn:lsid:marinespecies.org:taxname:299942	Bathypterois mediterraneus	
BIOFUN1_EF1M1_3	BF1M1	urn:lsid:marinespecies.org:taxname:280313	Coelorinchus mediterraneus	
BIOFUN1_EF1M1_4	BF1M1	urn:lsid:marinespecies.org:taxname:105812	Galeus melastomus	
BIOFUN1_EF1M1_5	BF1M1	urn:lsid:marinespecies.org:taxname:126495	Lepidion lepidion	
BIOFUN1_EF1M1_6	BF1M1	urn:lsid:marinespecies.org:taxname:126497	Mora moro	

	А	1		с	D	E
1	id 🝸	occurrenc	ID 🝸	measurementType 📼	measurementTypeID =	measurementValue \Xi
2	BF1M1	BIOFUN1_	BF1M1_1	DbservedIndividualCount	/P01/current/OCOUNT01	26
400	BF1M1	BIOFUN1_	BF1M1_1	Abundance	n/P01/current/SDBIOL02	0.000329114
816	BF1M1	BIOFUN1	BF1M1_1	Vet Weight Biomass	n/P01/current/SDBIOL05	0.091139241
1314	BF1A16			Gear	1/Q01/current/Q0100002	Agassiz dredge
1343	BF1A01			sampling net mesh size	1/Q01/current/Q0100015	12