Introduction to

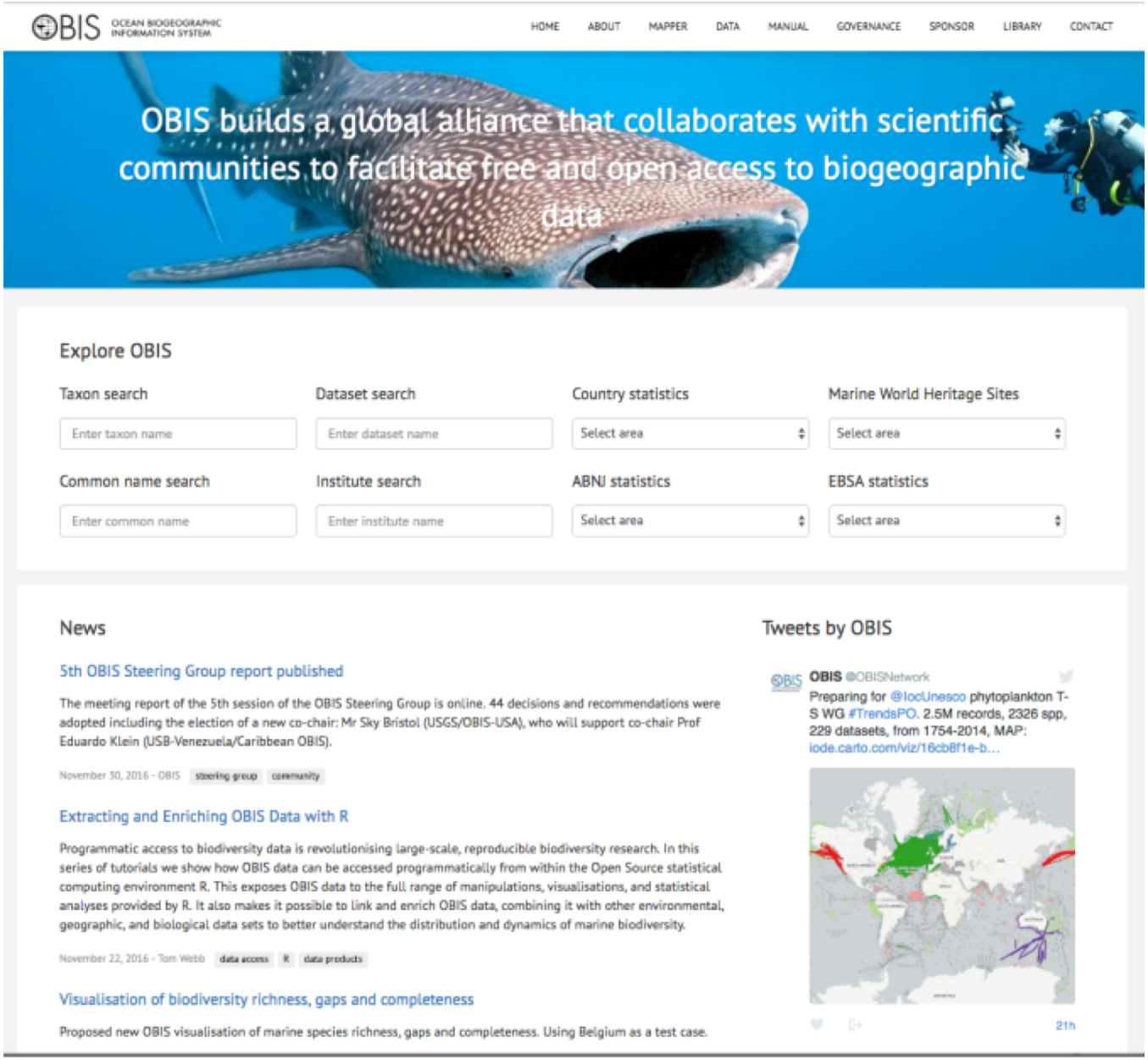
OBIS

OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

Abdolvahab (Vahab) Maghsoudlou
PEGO-OBIS node manager

OTGA/INIOAS: Remote Sensing of Coral Reefs
Training Course
20 - 23 October 2019

OTGA/INIOAS - Marine Biogeographic Data processing using OBIS 23 - 26 September 2018, Tehran, Iran (Islamic Republic of)



https://portal.obis.org/
https://obis.org

OBIS

- is the world's largest open access, online repository of spatially referenced marine life data
- was established by the **Census of Marine Life program** (www.coml.org) since 2000.
- In June 2009, the 25th Session of the IOC Assembly decided through Resolution XXV-4 to adopt OBIS as part of IODE
- OBIS was one of the earliest Associate Members of the Global Biodiversity Information Facility (www.gbif.org) which publishes data on all species.

United Nations Educational, Scientific and Cultural Organization

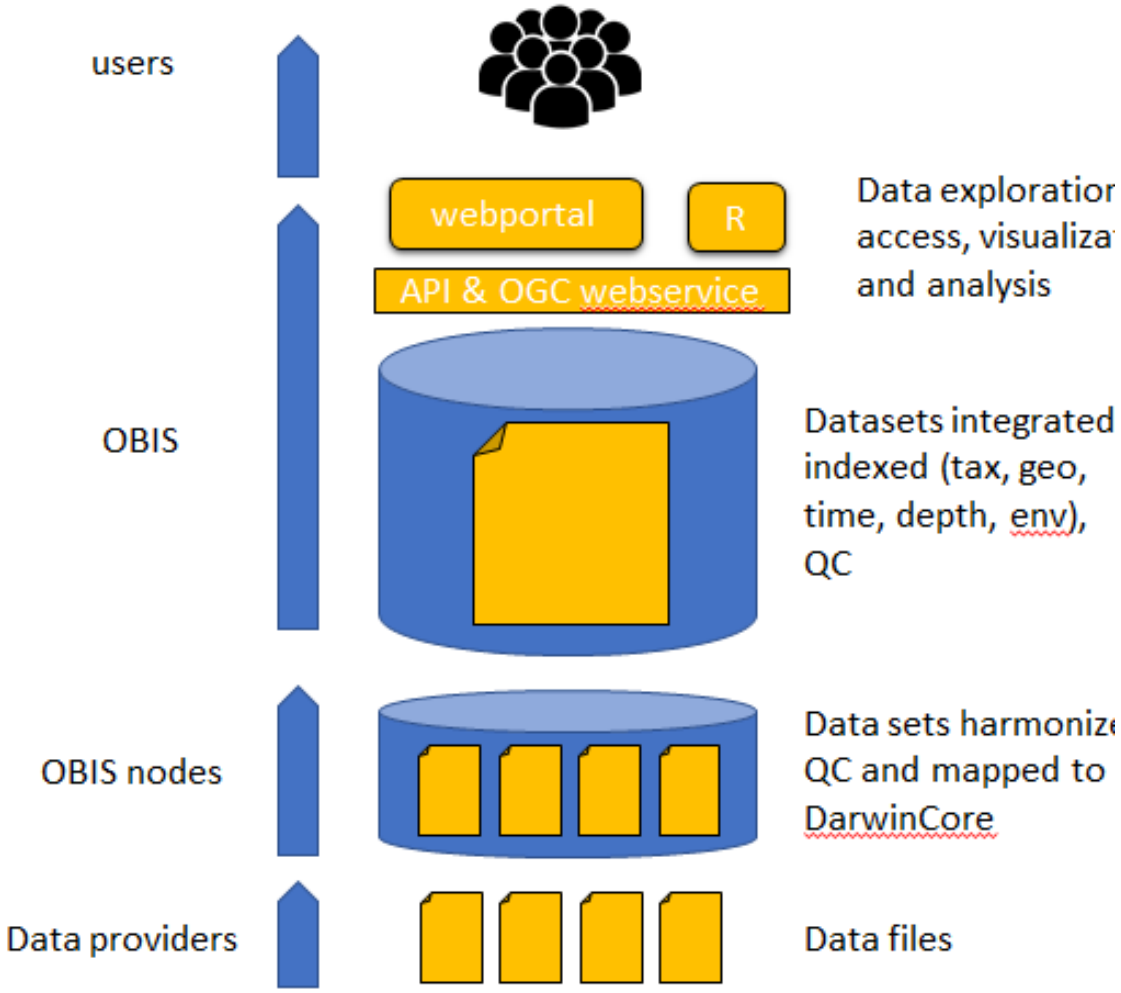
Intergovernmental Oceanographic Commission

GBIF Global Biodiversity Information Facility

CENSUS OF MARINE LIFE

OBIS international collaboration

OBIS process



>450 data providers in 56 countries

Scientists from 73 countries used OBIS in >1,000 research publications

ESP OBIS

Tropical and Subtropical Eastern South Pacific OBIS node

Website http://ron.udec.cl/

Pamela Hidalgo
pamelahidalgodiaz@gmail.com

Ruben Escribano
rescribano@udec.cl

EUROBIS

European OBIS node

Website http://www.eurobis.org

Leen Vandepitte
leen.vandepitte@vliz.be

Joana Beja
joana.beja@vliz.be

INDOBIS

Indian OBIS node

OBIS CHINA

Website http://www.iobis.org.cn/index_e.htm

Kuidong Xu
kxu@qdio.ac.cn

OBIS COLOMBIA

Website http://siam.invemar.org.co/

Martha Vides
martha.vides@invemar.org.co

Erika Montoya-Cadavid
erika.montoya@invemar.org.co

OBIS CPPS

CPPS - South Pacific OBIS node

Website http://cpps-int.org/index.php/2015-04-28-20-21-16/nodo-obis

Mónica Machuca

Esther Hughes
esthug@mba.ac.uk

OBIS USA

Website https://www.usgs.gov/obis-usa

Abby Benson
albenson@usgs.gov

Sky Bristol
sbristol@usgs.gov

PEGO OBIS

Persian Gulf and Gulf of Oman OBIS node

Website https://obis.org/node/066e070a-04ca-4cee-acb3-66379fe49d49

Abdolvahab Maghsoudlou
wahabblo@gmail.com

OBIS at technical level?

- Providing multiple datasets in a standard format for harvesting by iOBIS (*Darwin Core Archive format*)
- Having complete and good quality data and metadata for each dataset
- Regular updates to the datasets where possible
- IPT Server to serve the datasets (or use the IPT server of iOBIS to upload datasets)

OBIS uses the following standards:

- Darwin Core (species occurrence data)
- Ecological Metadata Language (dataset metadata)
- Darwin Core Archive including OBIS-ENV-DATA (sampling events and facts, species occurrences and measurements)

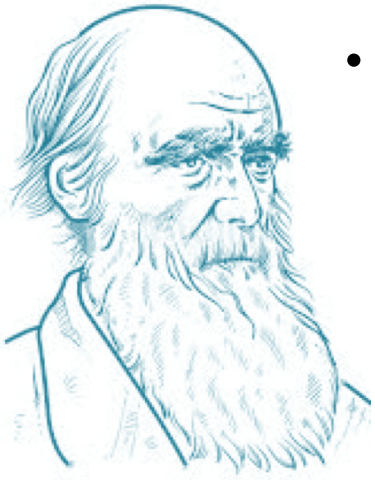
Ecological Metadata Language (EML)

OBIS (and GBIF) uses the Ecological Metadata Language (EML) as its metadata standard, which is specifically developed for the earth, environmental and ecological sciences.

. EML is implemented as XML. See more information on EML.

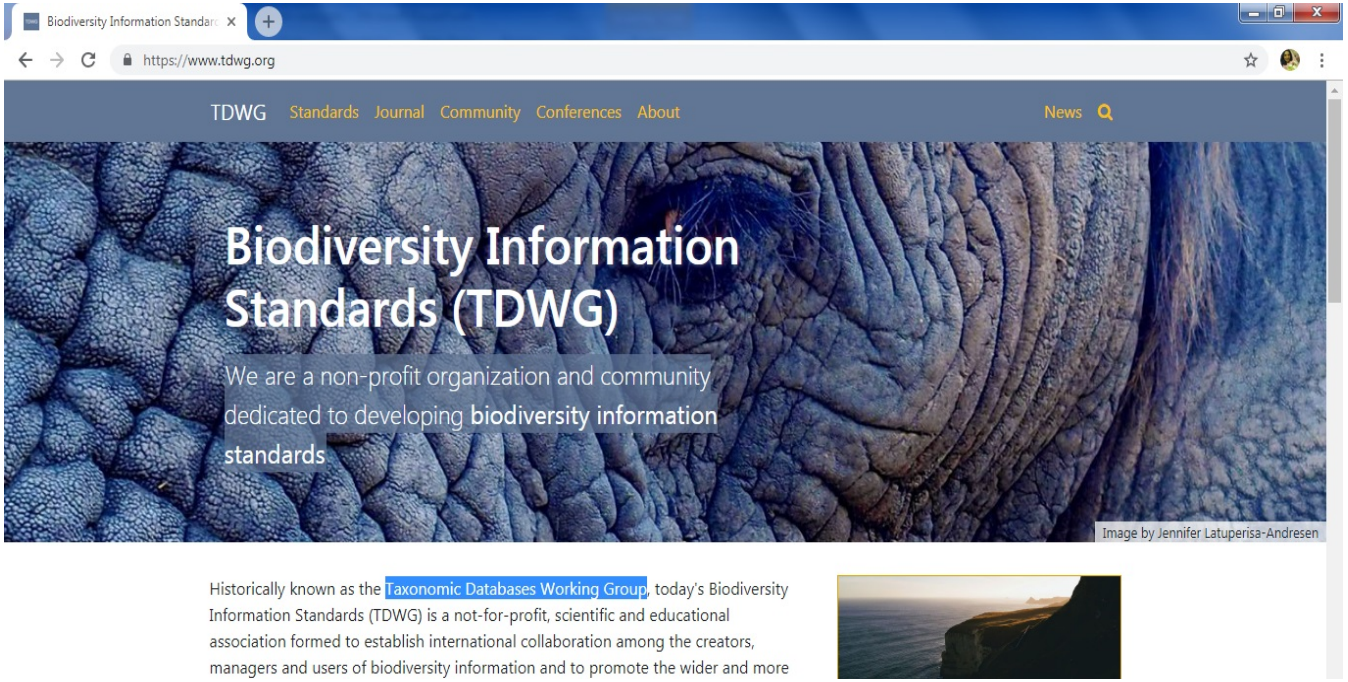
OBIS uses the GBIF EML profile (version 1.1).

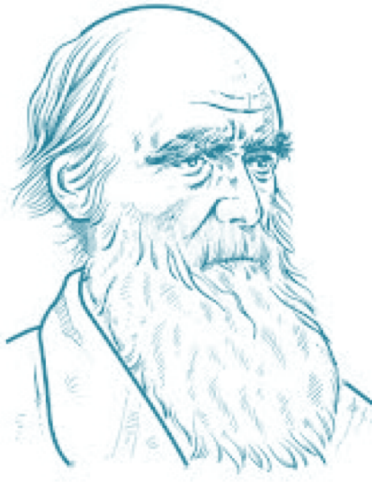
in case data providers use ISO19115/ISO19139,



- **Darwin Core** is a body of standards to create a **common language** for documenting and publishing data about species registers (field observations or preserved specimens in a collection).

- Started in 1999 and ratified as a standard in 2009 by the Dublin Core Metadata Initiative - DCMI and nowadays supported by the TDWG (Biodiversity Information Standards, formerly Taxonomic Databases Working Group).



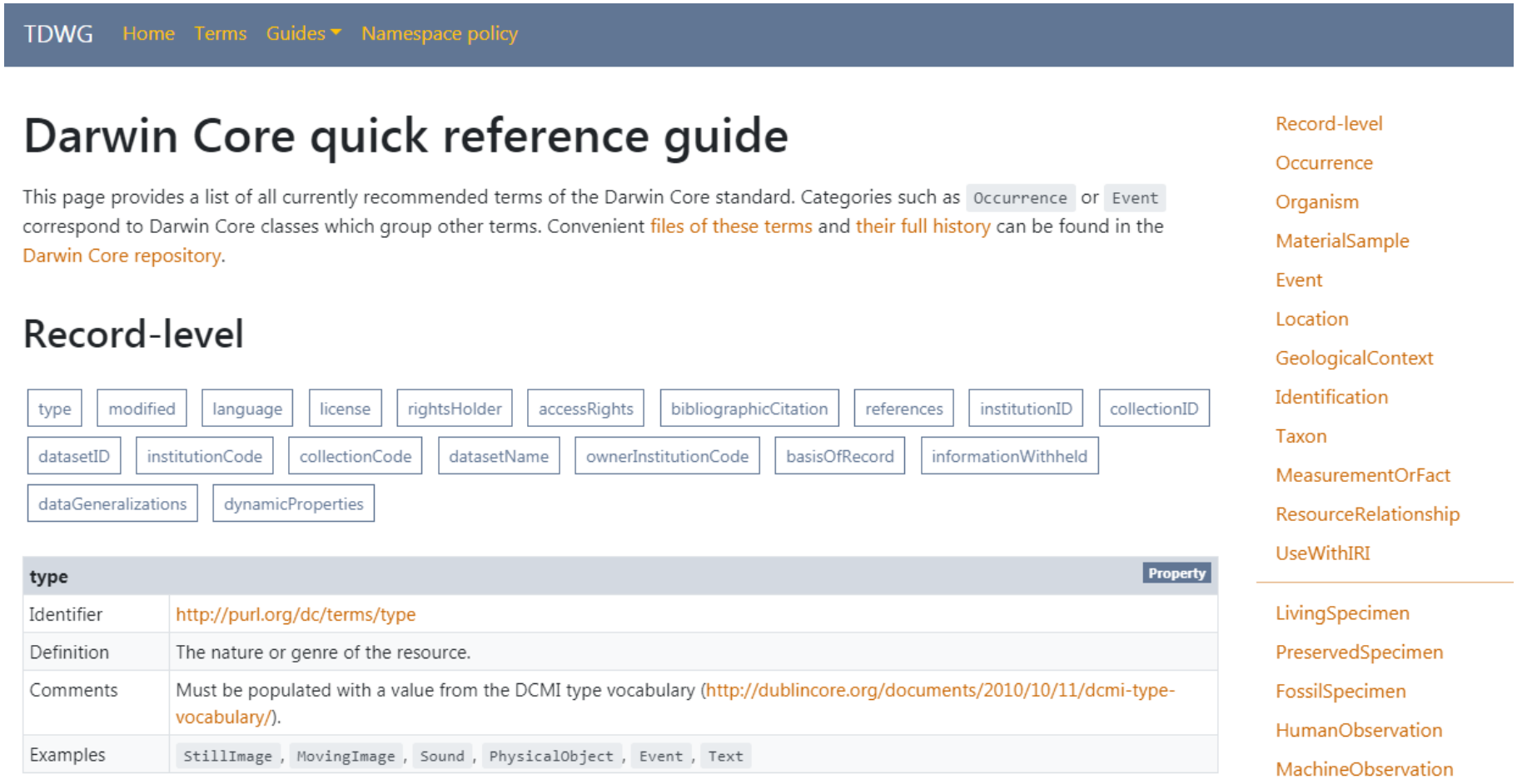


Darwin Core terms and Term Definitions

- **DwC terms** refer to the **column names** of your dataset.
- The complete Darwin Core template consists of 73 terms to cover many variables according to the record context. (**Full list is here :** http://rs.tdwg.org/dwc/terms/index.htm#Occurrence)

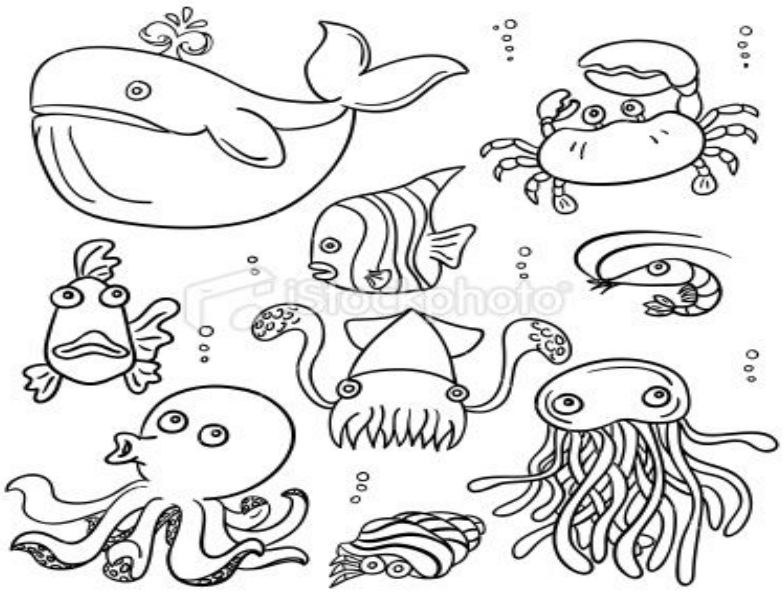
It provides stable terms and vocabularies for sharing biodiversity data.

https://dwc.tdwg.org/terms/



OBIS DATA STANDARDS

- **occurrenceID**
- **eventDate**
- **decimalLongitude** and **decimalLatitude**
- **scientificName**
- **scientificNameID**
- **occurrenceStatus**
  - present or absent
- **basisOfRecord (**DwC-A required term)
  - PreservedSpecimen, FossilSpecimen, LivingSpecimen, HumanObservation, MachineObservation



Darwin Core Archive (DwC- A)

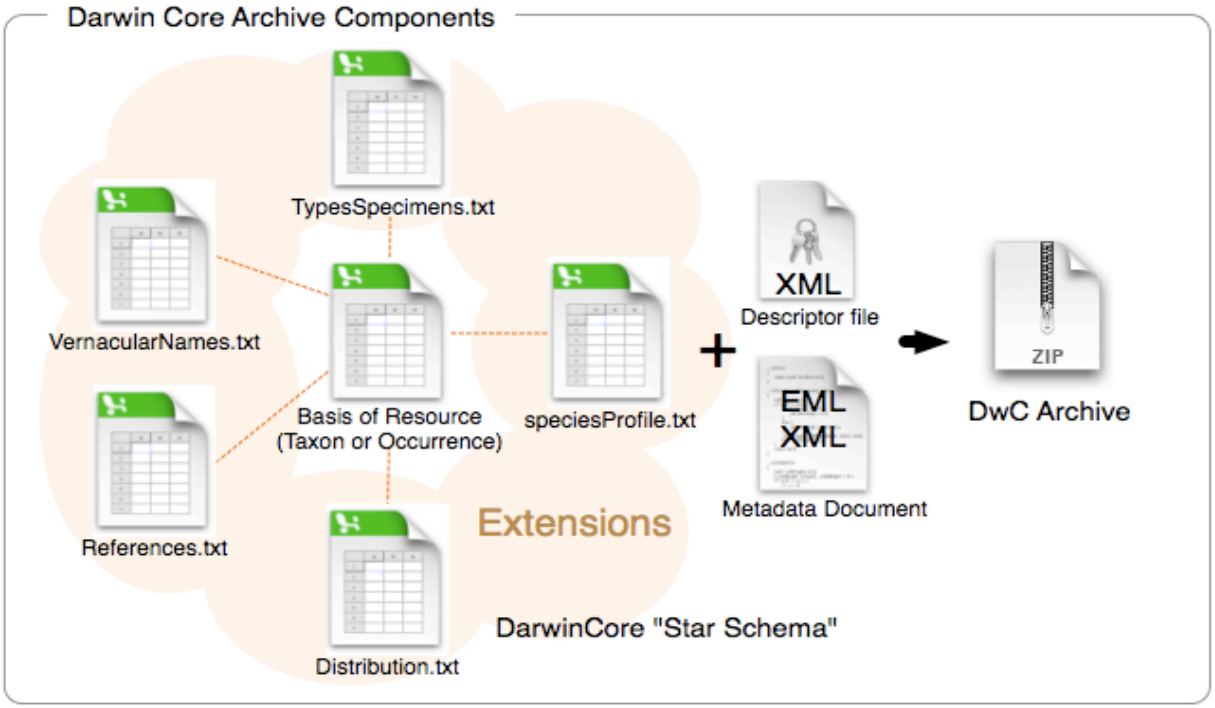
- is a standard for **publishing** biodiversity data using the Darwin Core format.
- **It is the preferred format for publishing data in OBIS and GBIF**

Darwin Core archives contain text files which are logically arranged in a star schema. This means that there is one **core file** and (optionally) multiple **extensions files**.

For example:

➔The species occurrence recorded in a research field trip = **core file.**

➔Environmental data, sampling methodology, etc. = **extension file.**



In 23 session of the IOC Committee for IODE, March 2015

OBIS-ENV-DATA

EXPANDING OBIS BEYOND OCCURRENCE DATA

As part of the IODE pilot project: Expanding OBIS with environmental data OBIS-ENV-DATA, OBIS introduced a customized ExtendedMeasurementOrFact Extension or eMoF, which extends GBIF's DwC MeasurementOrFact Extension with 4 new terms: **occurrenceID, measurementTypeID, measurementValueID and measurementUnitID**.

Extended MEASUREMENTS OR FACTS (eMoF)

- **ID**: the identifier used by DwC-A standard to link the eMoF to the Core file.
- **occurrenceID** (new): identifier to link the eMoF with the occurrence extension.
- **measurementType**: The nature of the measurement, fact, characteristic, or assertion.
  - **measurementTypeID** (new): An identifier for the measurementType (global unique identifier, URI)
- **measurementValue**: The value of the measurement, fact, characteristic, or assertion.
  - **measurementValueID** (new): An identifier for facts stored in the column measurementValue (global unique identifier, URI)
- **measurementAccuracy**: The description of the potential error associated with the measurementValue.
- **measurementUnit**: The value of the measurement, fact, characteristic, or assertion.
  - **measurementUnitID** (new): An identifier for the measurementUnit (global unique identifier, URI)
- **measurementDeterminedDate**: The date on which the MeasurementOrFact was made.
- **measurementDeterminedBy**: A list (concatenated and separated) of names of people, groups, or organizations who determined the value of the MeasurementOrFact.
- **measurementMethod**: A description of or reference to (publication, URI) the method or protocol used to determine the measurement, fact, characteristic, or assertion.
- **measurementRemarks**: Comments or notes accompanying the MeasurementOrFact.

Measurements or facts Vocabulary

The MoF terms: measurementType, measurementValue and measurementUnit are completely unconstrained and can be populated with free text annotation.

OBIS uses the **controlled vocabulary** developed and maintained by the British Oceanographic Data Center (BODC), and made available through the NERC Vocabulary server: https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/.

measurementTypeID

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

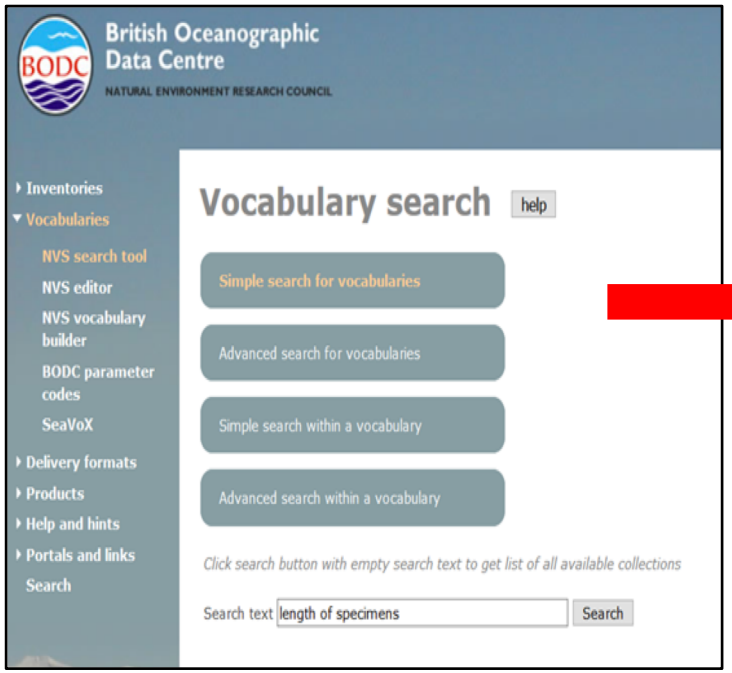
measurementValueID

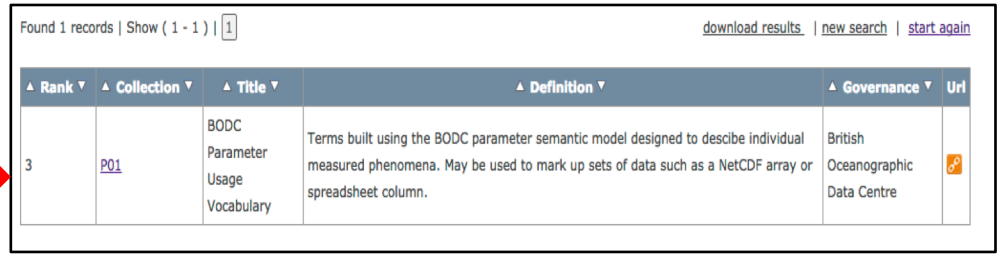
- Sampling instruments and sensors (SeaVoX Device Catalogue): http://vocab.nerc.ac.uk/collection/L22/current
- Sampling instrument categories (SeaDataNet device categories): http://vocab.nerc.ac.uk/collection/L05/current
- Vessels (ICES Platform Codes): http://vocab.nerc.ac.uk/collection/C17/current
- Lifestage: http://vocab.nerc.ac.uk/collection/S11/current/
- DOIs of papers or manuals on the sampling protocol used, published e.g. on IOC's Ocean Best Practices repository, for example: http://hdl.handle.net/11329/304

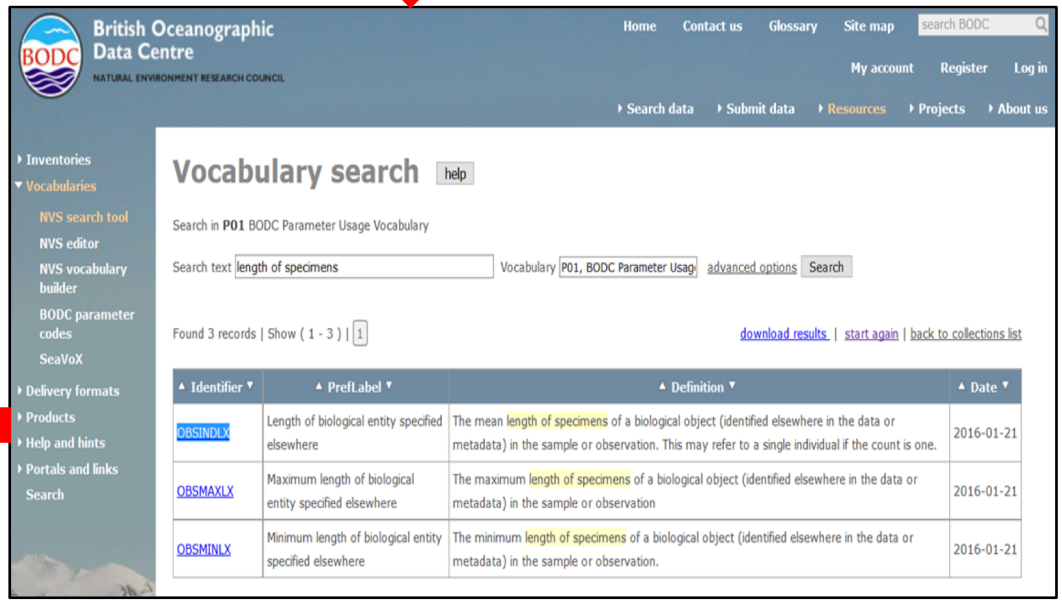
MeasurementUnitID

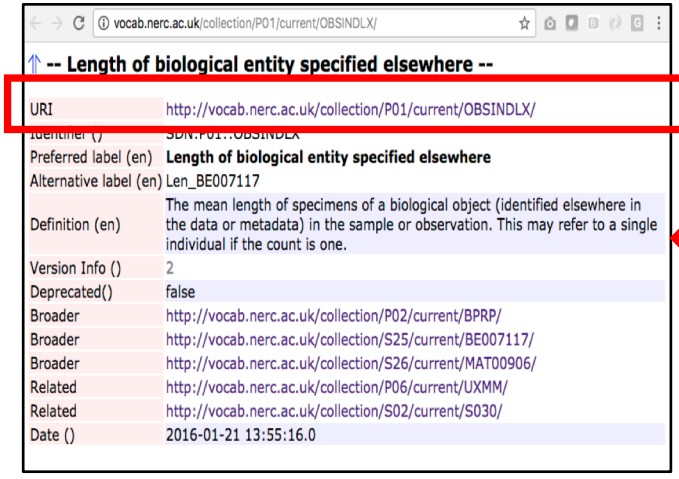
- Units: http://vocab.nerc.ac.uk/collection/P06/current

https://www.bodc.ac.uk/resources/vocabularies/vocabulary_search/.









Data Standardization

⇑ -- Length of biological entity specified elsewhere --

URI	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX/
Identifier ()	SDN:P01::OBSINDLX
Preferred label (en)	**Length of biological entity specified elsewhere**
Alternative label (en)	Len_BE007117
Definition (en)	The mean length of specimens of a biological object (identified elsewhere in the data or metadata) in the sample or observation. This may refer to a single individual if the count is one
Version Info ()	2
Deprecated()	false
Broader	http://vocab.nerc.ac.uk/collection/P02/current/BPRP/
Broader	http://vocab.nerc.ac.uk/collection/S25/current/BE007117/
Broader	http://vocab.nerc.ac.uk/collection/S26/current/MAT00906/
Related	http://vocab.nerc.ac.uk/collection/P06/current/UXMM/
Related	http://vocab.nerc.ac.uk/collection/S02/current/S030/
Date ()	2016-01-21 13:55:16.0

Controlled vocabulary
Persistent identifiers (URIs)

MeasurementUnitID

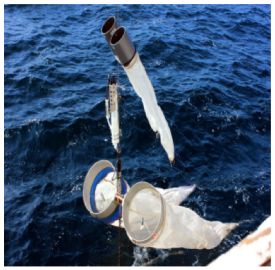
MeasurementType	**MeasurementTypeID**
Body length	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length (mm)	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
length_in_m	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
Length of specimen	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX
…	

OBIS accepts 2 types of core files: **Occurrence Core** and **Event Core**

01 Data schema / structure

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...)



Depending on the types of data you can organize your dataset in three different ways:

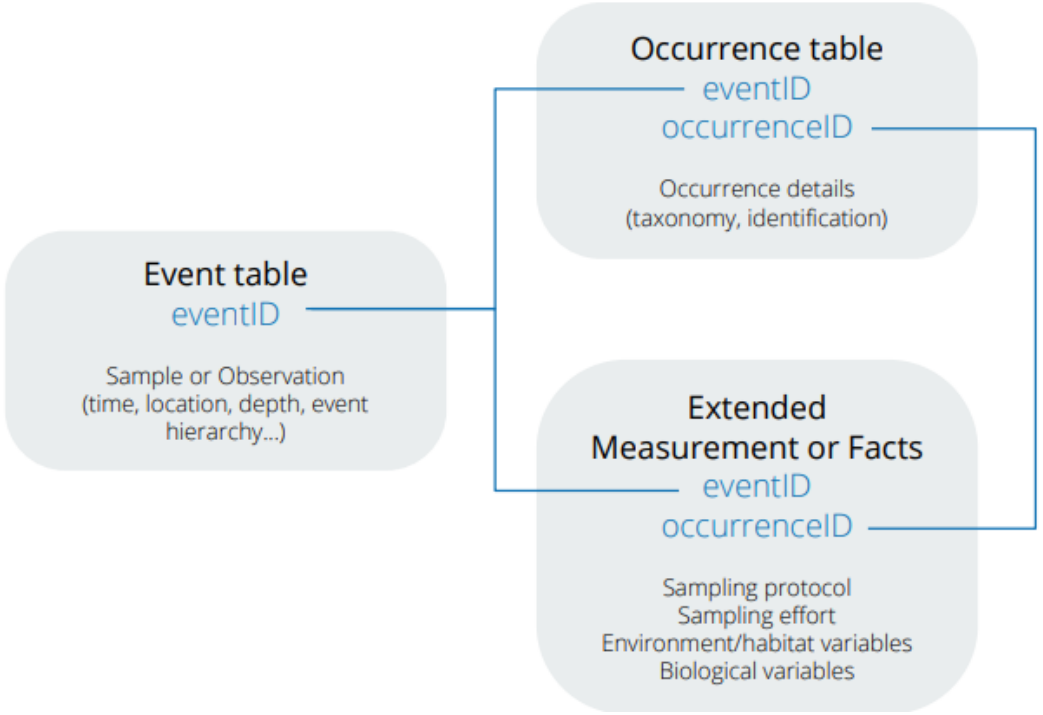
- **Case 1:** One OCCURRENCE file (Occurrence Core). Entering only the details of the occurrences.
- **Case 2:** Two sets of files: OCCURRENCE (Occurrence Core) + MEASUREMENTS (Measurement or Fact extension), including the measurements or facts made on each specimen or sample (e.g. size, abundance, wet weight, life stage, etc).

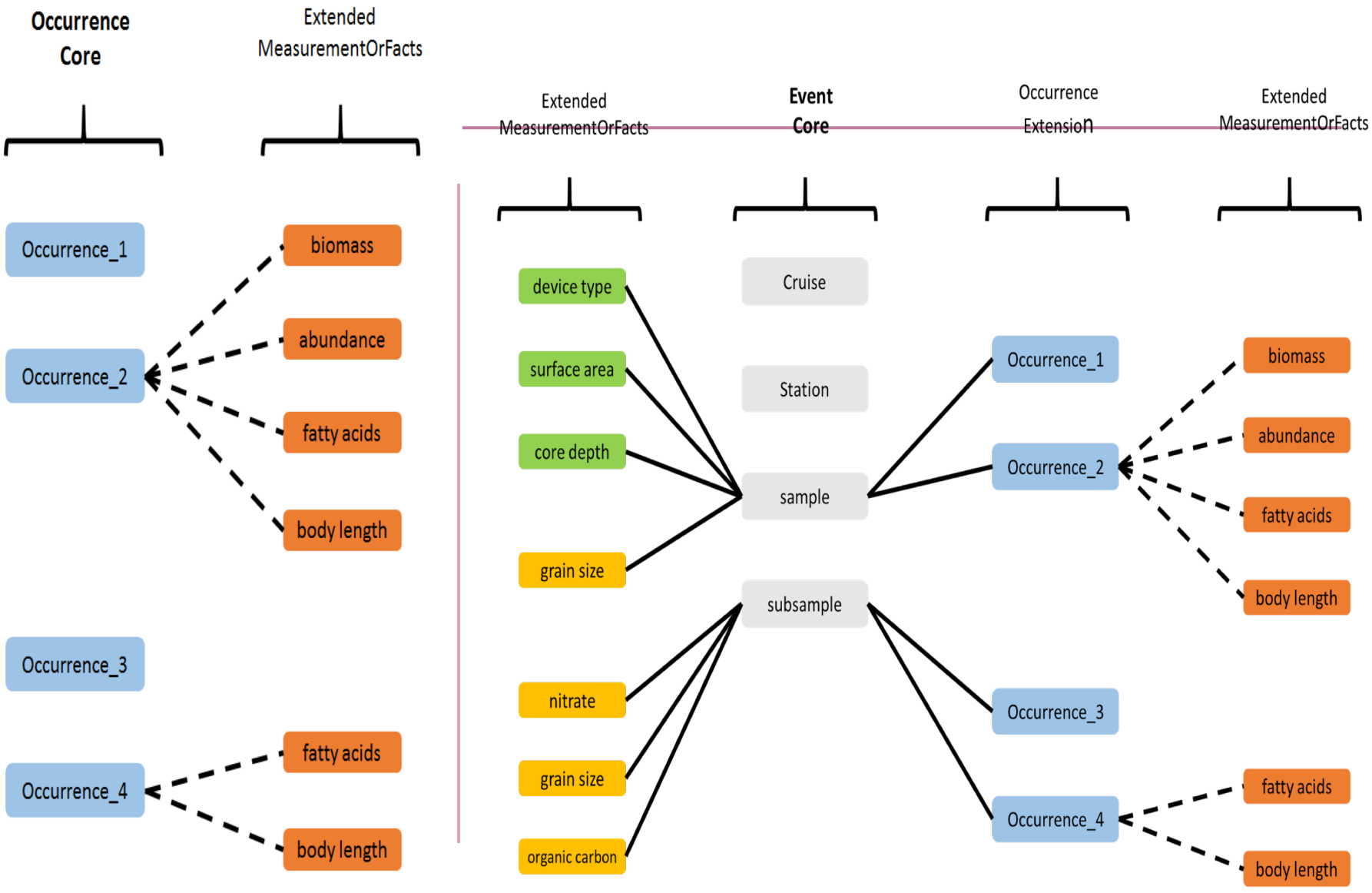
- **Case 3:** Three sets of files: EVENT (Event Core), the sampling event details (position, time, depth of each site, station, sample etc).+ OCCURRENCE (Occurrence extension) + MEASUREMENTS (Measurement or Fact extension), including environmental data, the measurements made on each specimen (e.g. size, wet weight), sampling facts.

01 Data schema / structure

The three tables are related via the eventID and the occurrenceID.

The eMoF Extension is used in combination with the Event Core and the Occurrence Extension to capture both abiotic measurements and biotic measurements. The occurrenceID is used to link biotic measurements in the eMoF Extension with the the Occurrence Extension and the eventID links the eMoF to the Event Core.





Full lines: eventID links; Dashed lines: occurrenceID links.

When we 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.

scientificName	scientificNameID	occurrenceID	eventDate	decimalLatitude	decimalLongitude	occurrenceStatus	basisOfRecord
Arca zebra	urn:lsid:marinespecies.org:taxname:420713	MCNUSB_001	1999-01-01	10.7413	-63.8791	Present	PreservedSpecimen
Perna viridis	urn:lsid:marinespecies.org:taxname:367822	MCNUSB_002	1999-01-01	10.7413	-63.8791	Present	PreservedSpecimen
Phyllonotus pomum	urn:lsid:marinespecies.org:taxname:419944	MCNUSB_003	1999-01-01	10.7413	-63.8791	Present	PreservedSpecimen
Strombus pugilis	urn:lsid:marinespecies.org:taxname:419695	MCNUSB_047	1999-01-01	10.8737	-63.8805	Present	PreservedSpecimen
Trachycardium	urn:lsid:marinespecies.org:taxname:203976	MCNUSB_075	1999-01-01	10.8477	-68.2424	Present	PreservedSpecimen
Chione cancellata	urn:lsid:marinespecies.org:taxname:397040	MCNUSB_006	1999-01-01	10.6886	-63.8514	Present	PreservedSpecimen
Atrina seminuda	urn:lsid:marinespecies.org:taxname:420740	MCNUSB_007	1999-01-01	10.6886	-63.8514	Present	PreservedSpecimen
Lyropecten	urn:lsid:marinespecies.org:taxname:203879	MCNUSB_004	1999-01-01	10.7413	-63.8791	Present	PreservedSpecimen

Event Core

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.
- Event Core should be used in combination with the Occurrence Extension and the ExtendedMeasurementOrFact Extension.

Taxon terms

- **scientificName** (required)
- contain the originally recorded scientific name, even if the name is currently a synomym. This is necessary to be able to track back records to the original dataset.
  - lowest level taxonomic rank that can be determined (but higher ranks, such as genus, family, order, class etc are also acceptable)
  - no identification qualifiers (cf., aff.), see identificationQualifier
  - **OBIS recommended practice: no authorship**
- **scientificNameID**
  - WoRMS LSID, no matter if the taxonomic status is accepted or not. Double check if authority and kingdom are correct
  - urn:lsid:marinespecies.org:taxname:141433

scientificName	scientificNameID	identificationQualifier
Peltodoris atromaculata	urn:lsid:marinespecies.org:taxname:509315	
Peltodoris	urn:lsid:marinespecies.org:taxname:225423	cf. atromaculata

OCCURRENCE

- Terms
  - basisOfRecord
  - **occurrenceStatus**
  - **occurrenceID**
    - gobally unique
    - urn:catalog:[institutionCode]:[collectionCode]:[catalogNumber or autonumber in the absence of a catalogNumber]
    - ccurrenceID is also necessary for datasets in the OBIS-ENV-DATA format.

OCCURRENCE

- Terms
- basisOfRecord
- **PreservedSpecimen**: when specimen is deposited in a collection (please add institutionCode, collectionCode and CatalogNumber)
- **FossilSpecimen**: important to distinguish collection date from geological time zone
- **LivingSpecimen**: an intentionally kept/cultivated living specimen e.g. in culture collectionan aquarium or
- **HumanObservation**: e.g. bird sighting, benthic sample but specimens were discarded after counting
- **MachineObservation**: sensors, e.g. DNA sequencers, image recognition
- **occurrenceStatus**
- Present or Absent (individualCount = 0)
- **occurrenceStatus:** a statement about the presence or absence of a Taxon at a Location. Use “Present”or “Absent”.

LOCATION

- Terms
  - **decimalLongitude**, **decimalLatitude**
  - **coordinateUncertaintyInMeters**
  - geodeticDatum
    - OBIS recommended practice: **EPSG:4326**
  - **locationID**
    - for example MRGID from http://www.marineregions.org
  - **minimumDepthInMeters**, **maximumDepthInMeters**

LOCATION

Before conversion to decimals	In decimal format
18°30’25’’N – 5°15’E	18.51; 5.25
54,23N – 16,5S	54.23 ; -16.5

decimalLatitude:the geographic latitude for the occurrence register. Must be in **decimal degrees**. For example: 12.2354 (for the Northern hemisphere); **-**12.2354 (for the **S**outhern hemisphere).

decimalLongitude:the geographic longitude for the occurrence register. Must be in decimal degrees. For example: 68.357 (for the eastern hemisphere); **-**68.357 (for **W**estern hemisphere).

If the locality is known but not the exact coordinates you could search in geocoding services: Marine Regions or Google Maps

- **locationID:** an identifier for the set of location information. You may find this information using the Marine Regions Searching tool: you will find the MRGID. Copy and paste it in your data set excel file.

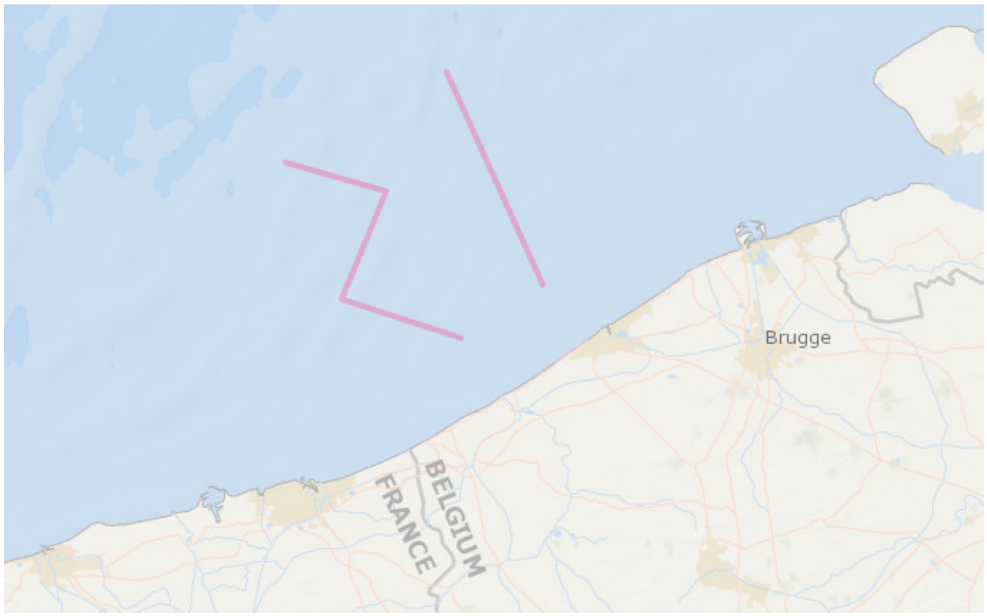
locationID	continent	waterBody
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters
http://marineregions.org/mrgid/21190	South America	Open waters

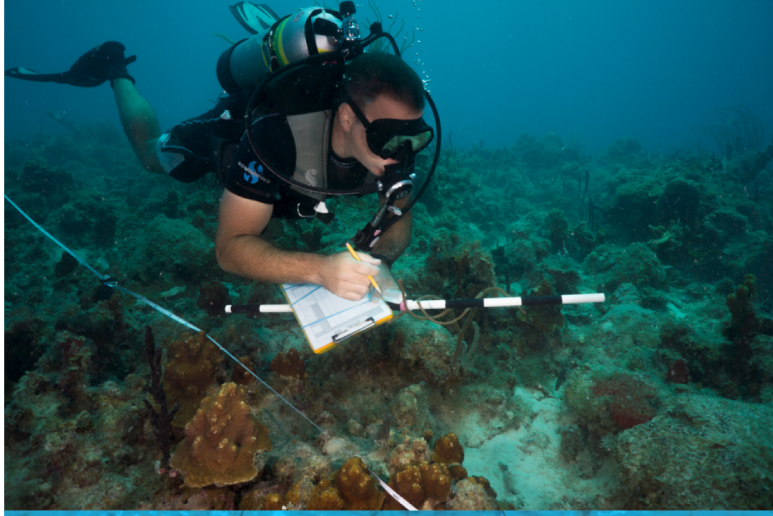
LOCATION

footprintWKT: transects -

```
LINESTRING (2.80151 51.28597,
            2.61749 51.53950)

LINESTRING (2.64496 51.22237,
2.41699 51.26879, 2.50214 51.39749,
            2.30988 51.43175)
```





LOCATION

footprintWKT: polygons -

```
POLYGON ((10.65674 42.77928,
 10.50018 42.77121, 10.43152
42.62183, 10.75836 42.38087,
 11.05225 42.48628, 10.91492
42.70262, 10.65674 42.77928))
```

OBIS
OCEAN BIOGEOGRAPHIC
INFORMATION SYSTEM

LOCATION

Examples -

locality	locationID	decimal Longitude	decimal Latitude	coordinate Uncertainty InMeters	minimum Depth InMeters	maximum Depth InMeters	footprint WKT
Ha Long Bay	MRGID:8897	107.1	20.9	26000			
station_115		2.6999	51.2219	50	30	30	
		2.5996	51.2765	8134	5	10	LINESTRING (2.53510 51.21549, 2.66418 51.33748)

TIME

- Terms
  - **eventDate**
  - verbatimEventDate (not recommended)
- ISO 8601
  - 1973-02-28T15:25:00
  - 1973-02-28 (yyyy-mm-dd)
  - 1973-02
  - 1973
  - 1973-02-28T15:25:00Z (UTC)
  - 1993-01-26T04:39+12/1993-01-26T05:48+12
  - 2015-023
  - 2014-W26

TIME

- **Not ok:**
  - 2015/07/11
  - 1915-6-9 0:00:00
  - 1995-7-0
  - 09-Dec-2009
  - 10-01-2013
  - 1:25
  - 00:18:00+0:00
  - Jan

IDENTIFIERS

Terms –

institutionCode –

institution who has custody over the collection or dataset –

collectionCode –

identifier for the collection or dataset (same for all records) –

catalogNumber –

unique key within the dataset –

occurrenceID –

gobally unique –

urn:catalog:[institutionCode]:[collectionCode]:[catalogNumber] –

recordNumber –

organismID –

IDENTIFIERS

Examples -

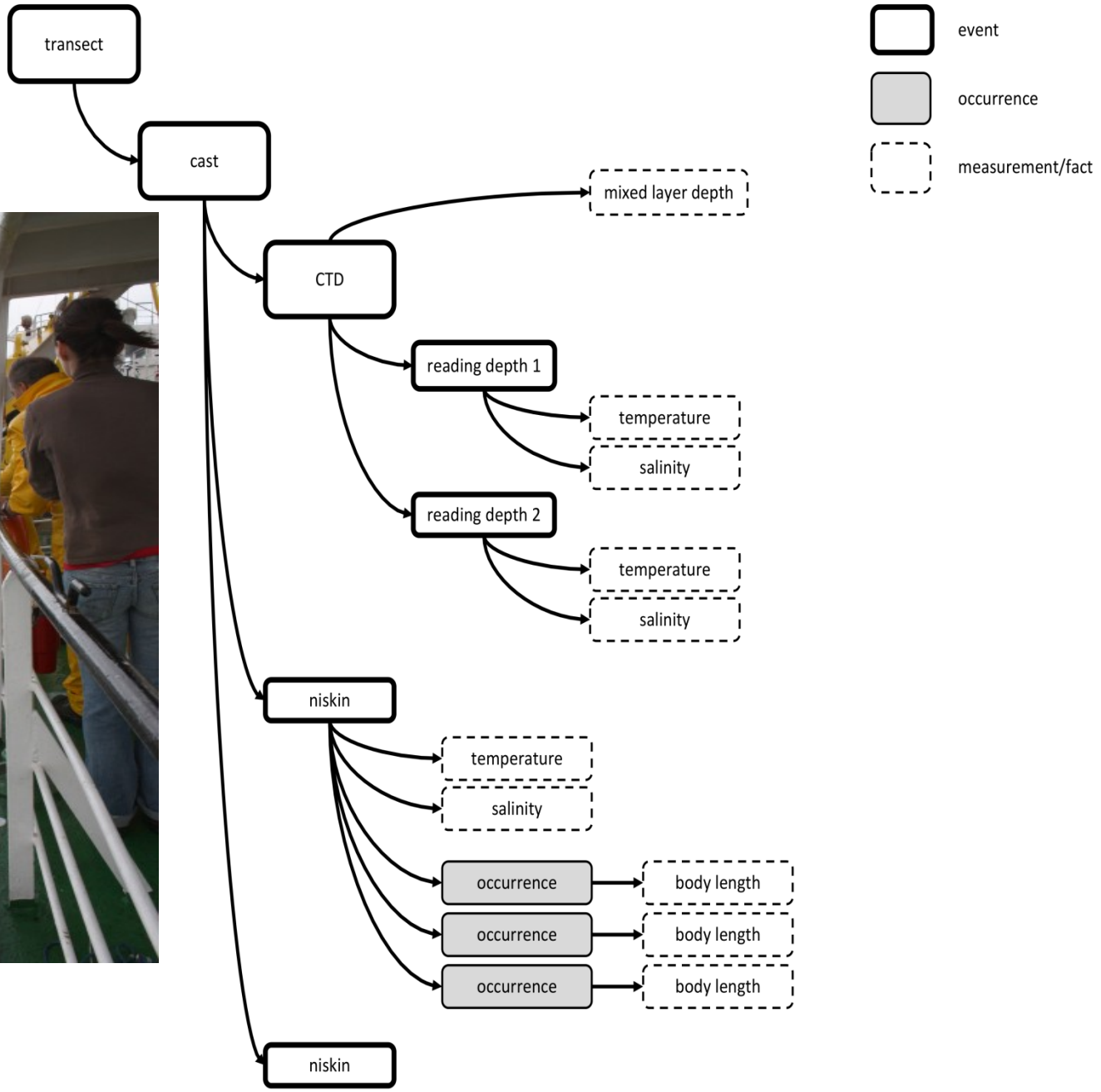
institutionCode	collectionCode	catalogNumber	occurrenceID
UGent	Macrobel	28125	urn:catalog:UGent:Macrobel:28125
ICES	DATRAS-EVHOE	865761	urn:catalog:ICES:DATRAS-EVHOE:865761

IDENTIFIERS

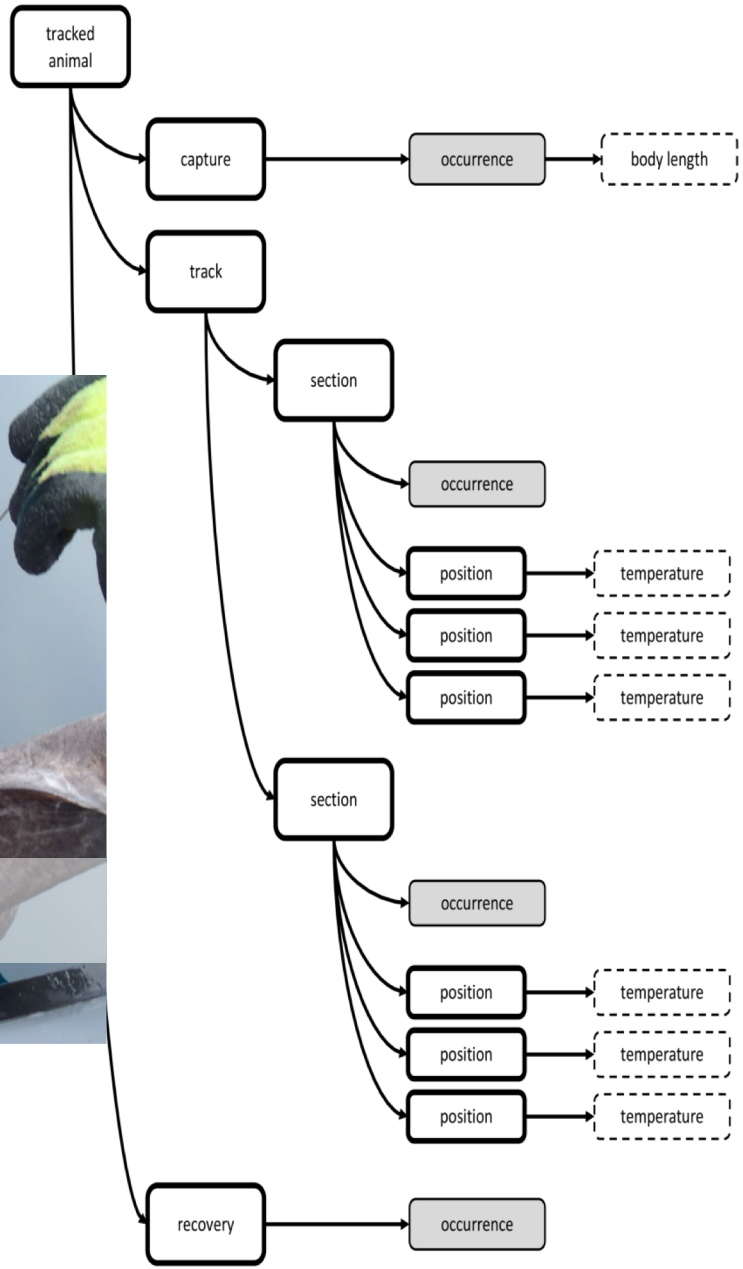
Example -

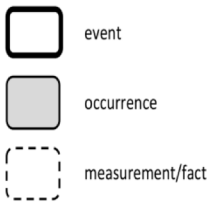
eventID	parentEventID	type	eventDate	decimalLongitude	decimalLatitude
cruise_1		cruise			
station_1	cruise_1	station		-12.0190	33.9069
grab_1	station_1	grab	2016-01-02T16:02		
grab_2	station_1	grab	2016-01-02T16:24		
subsample_1	grab_1	subsample			

eventID	scientificName
grab_1	Abra alba
grab_1	Lanice conchilega
subsample_1	Sabatieria armata

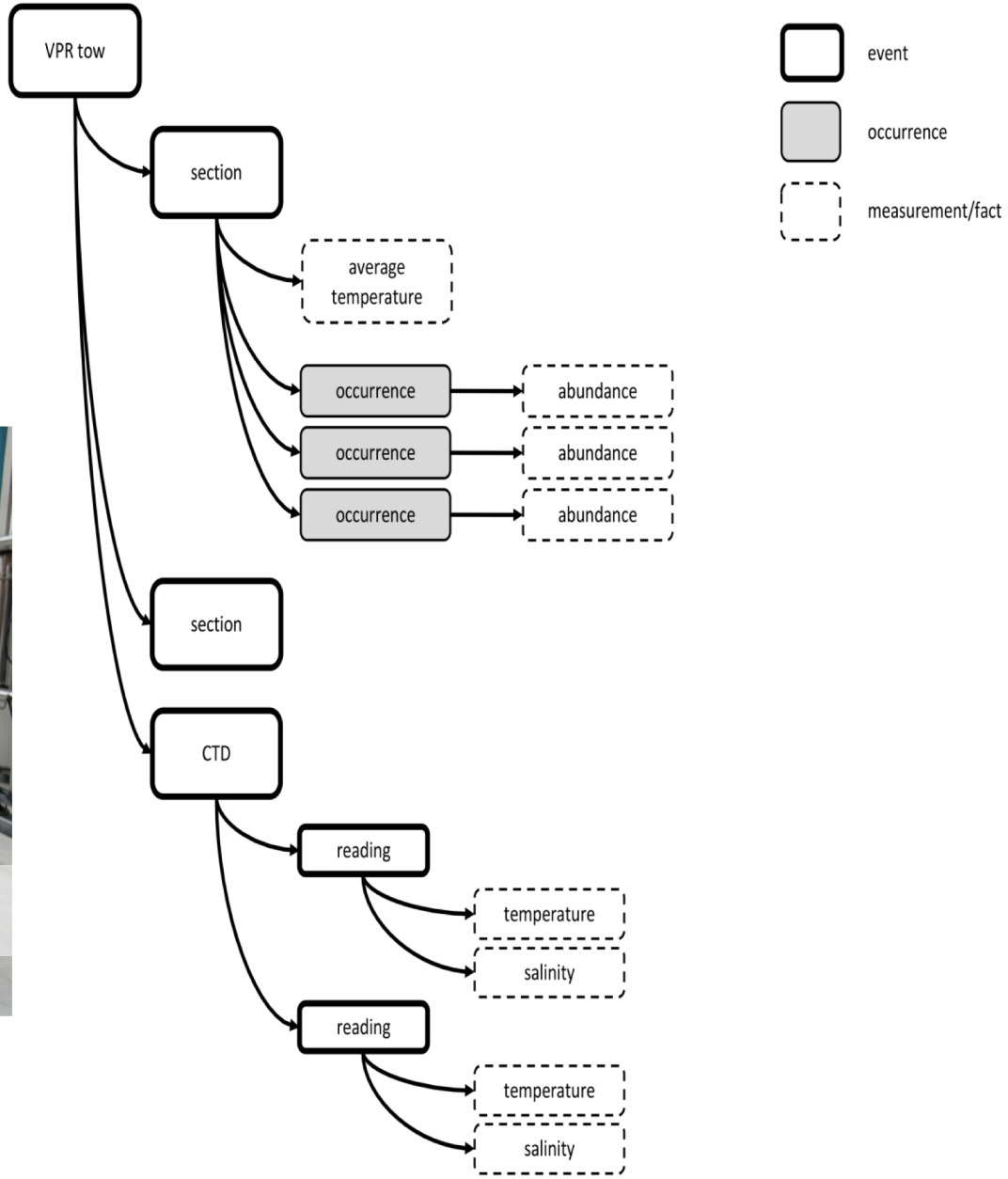

transect
cast
CTD
mixed layer depth
reading depth 1
temperature
salinity
reading depth 2
temperature
salinity
niskin
temperature
salinity
occurrence
body length
occurrence
body length
occurrence
body length
niskin
event
occurrence
measurement/fact

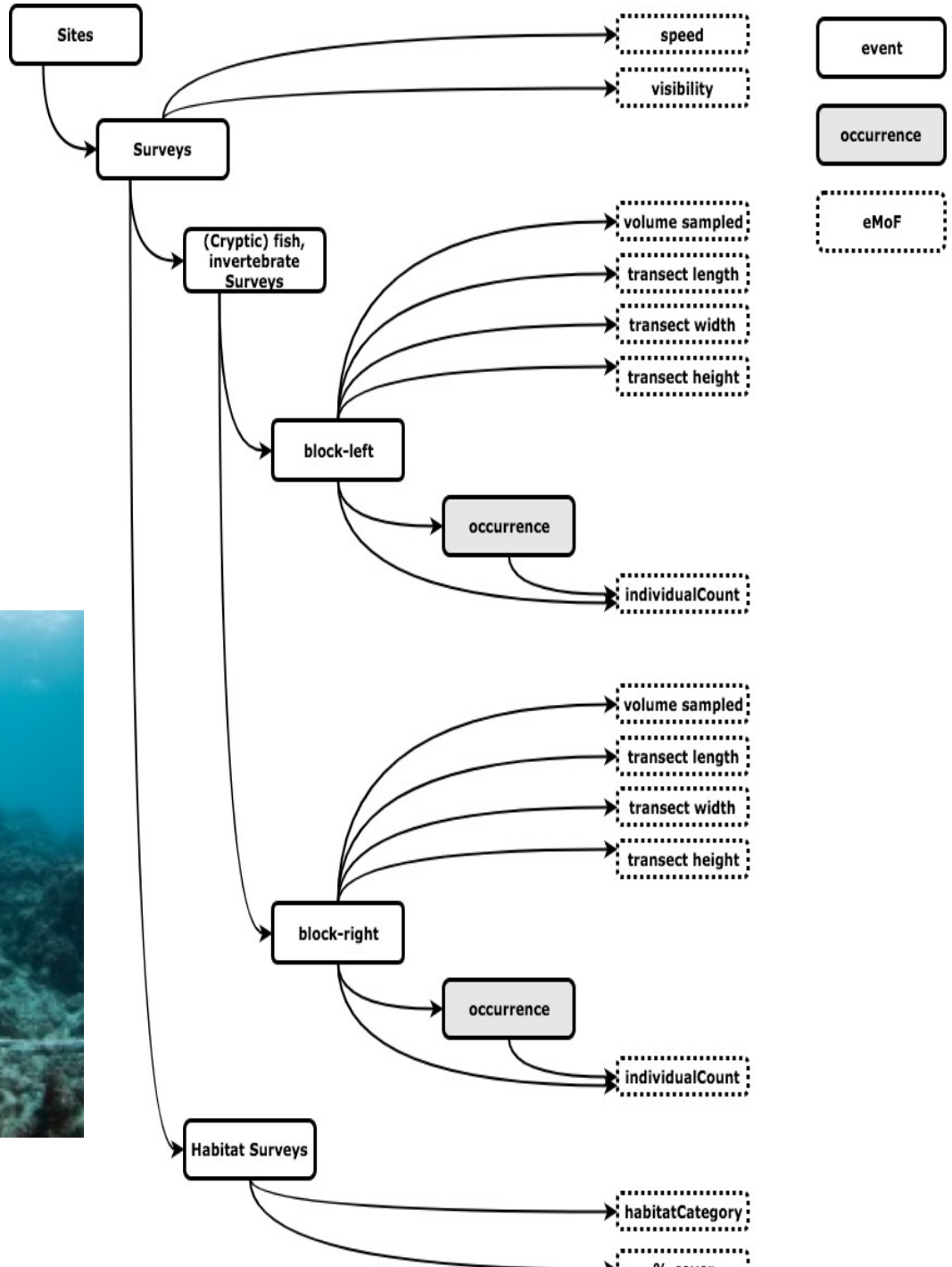

Tagged animals


tracked animal
capture
occurrence
body length
track
section
occurrence
position
temperature
position
temperature
position
temperature
section
occurrence
position
temperature
position
temperature
position
temperature
recovery
occurrence


event
occurrence
measurement/fact


VPR


VPR tow
section
average temperature
occurrence
abundance
occurrence
abundance
occurrence
abundance
section
CTD
reading
temperature
salinity
reading
temperature
salinity
event
occurrence
measurement/fact


Sites
Surveys
speed
visibility
(Cryptic) fish, invertebrate Surveys
volume sampled
transect length
transect width
transect height
block-left
occurrence
individualCount
volume sampled
transect length
transect width
transect height
block-right
occurrence
individualCount
Habitat Surveys
habitatCategory
event
occurrence
eMoF

Data is structured in 3 tables that are related to each other via the eventID and the occurrenceID:

Dataset example

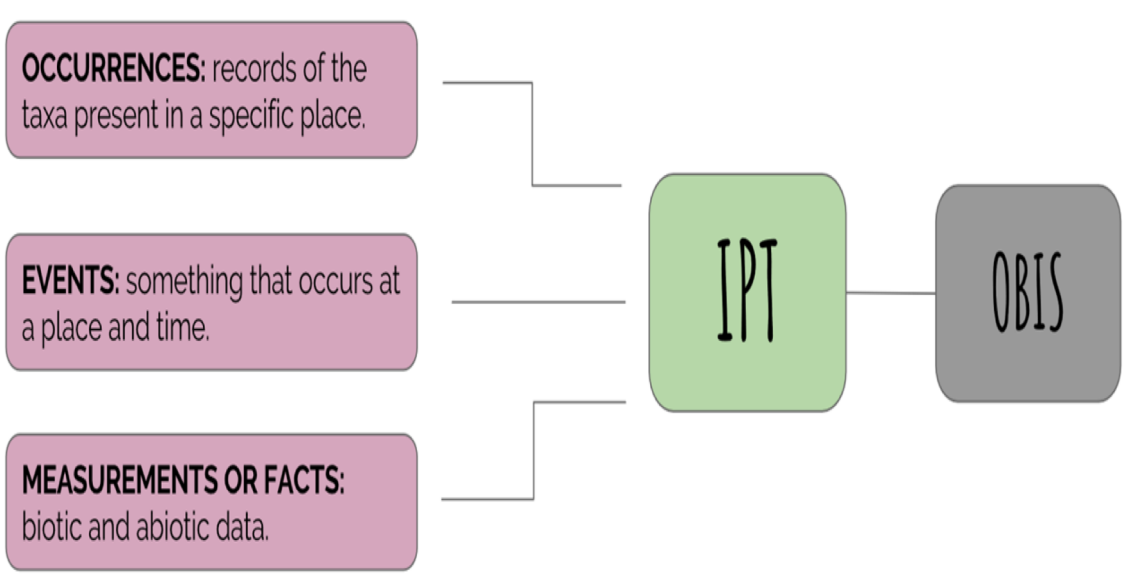
D	E	F	G	H	I	J	K
eventID	**parentEventID**	**eventDate**				**decimalLatitude**	**decimalLongitude**
BF1M1	BIOFUN1	30/05/2009				38.3881	1.8165
BF1M2	BIOFUN1	30/05/2009				38.3915	1.8125
BF1M3	BIOFUN1	01/06/2009				38.038	1.9027

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

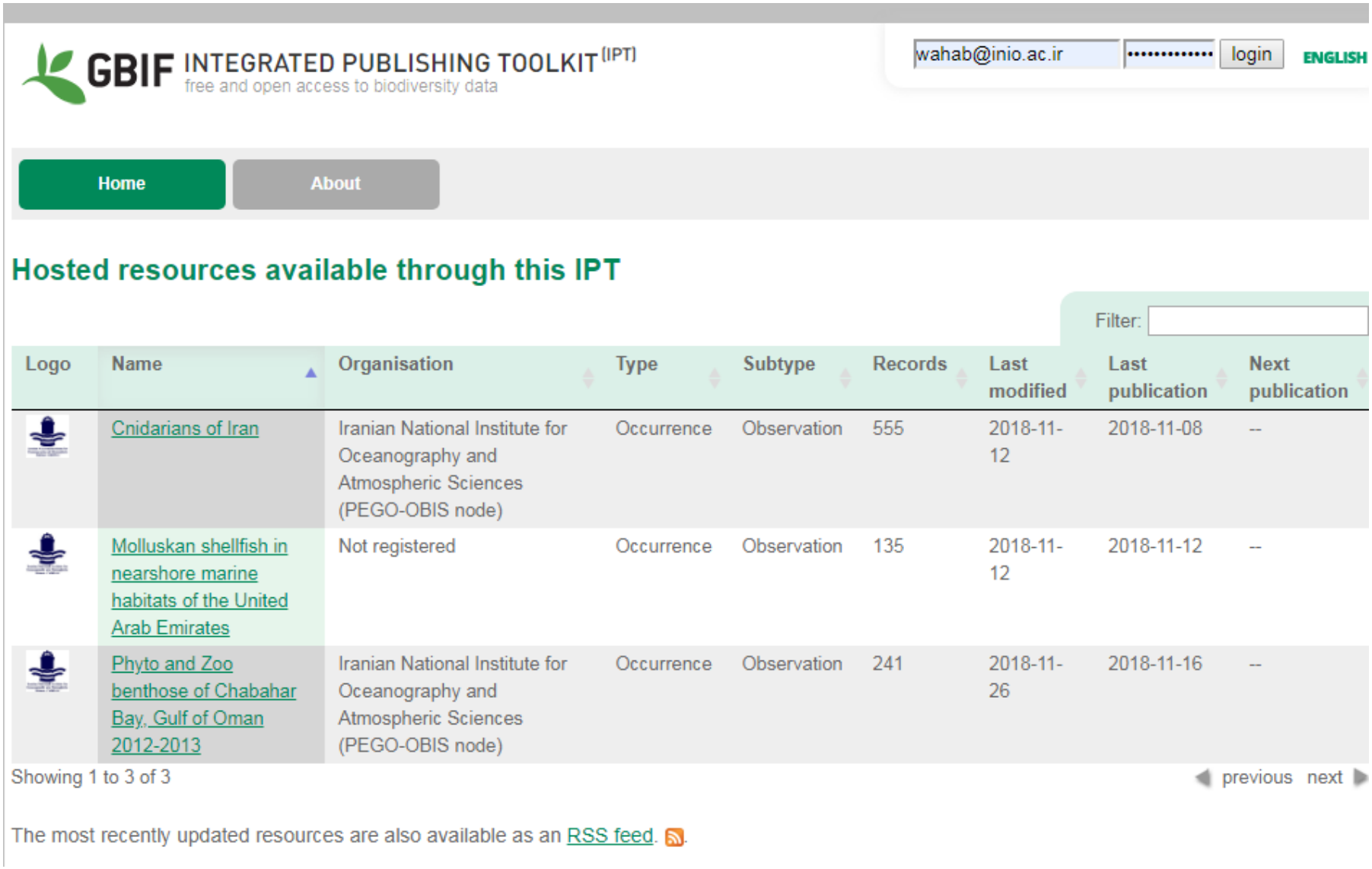
	A	B	C	D	E
1	**id**	**occurrenceID**	**measurementType**	**measurementTypeID**	**measurementValue**
2	BF1M1	BIOFUN1_BF1M1_1	ObservedIndividualCount	/P01/current/OCOUNT01	26
[illegible]	BF1M1	BIOFUN1_BF1M1_1	Abundance	n/P01/current/SDBIOL02	0.000329114
816	BF1M1	BIOFUN1_BF1M1_1	Wet Weight Biomass	n/P01/current/SDBIOL05	0.091139241
1314	BF1A16		Gear	/Q01/current/Q0100002	Agassiz dredge
1343	BF1A01		sampling net mesh size	/Q01/current/Q0100015	12

Publishing and sharing data

Datasets are published through IPT (the Integrated Publishing Toolkit). This tools allows you to add metadata (description of the dataset) and map the column names of each file with DarwinCore terms (see Module 3). Once published, OBIS can harvest the dataset and integrate it into the OBIS central database.



INIOAS IPT http://217.11.23.22:8080/ipt/



http://iobis.github.io/plotter/

https://mapper.obis.org/