

Parallel Publishing to GBIF and OBIS

Elizabeth Lawrence, Steve Formel, Dave Bloom
UNESCO IOC IODE Ocean Biodiversity Information System
e.lawrence@obis.org





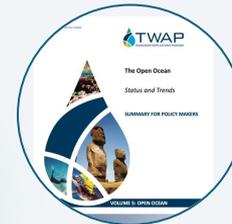
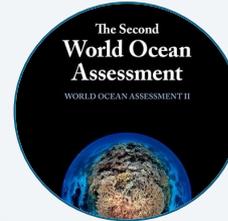
The world's largest open-access repository for **standardized marine biodiversity data**

36 Regional and thematic Nodes

Dedicated to be a **catalyst for marine science, conservation efforts, and policy-making**

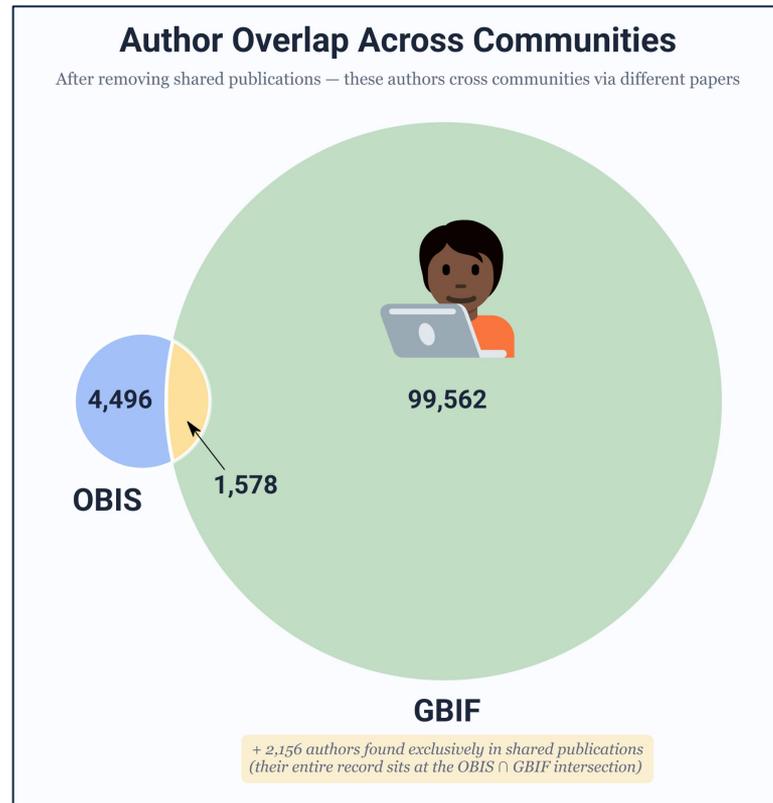
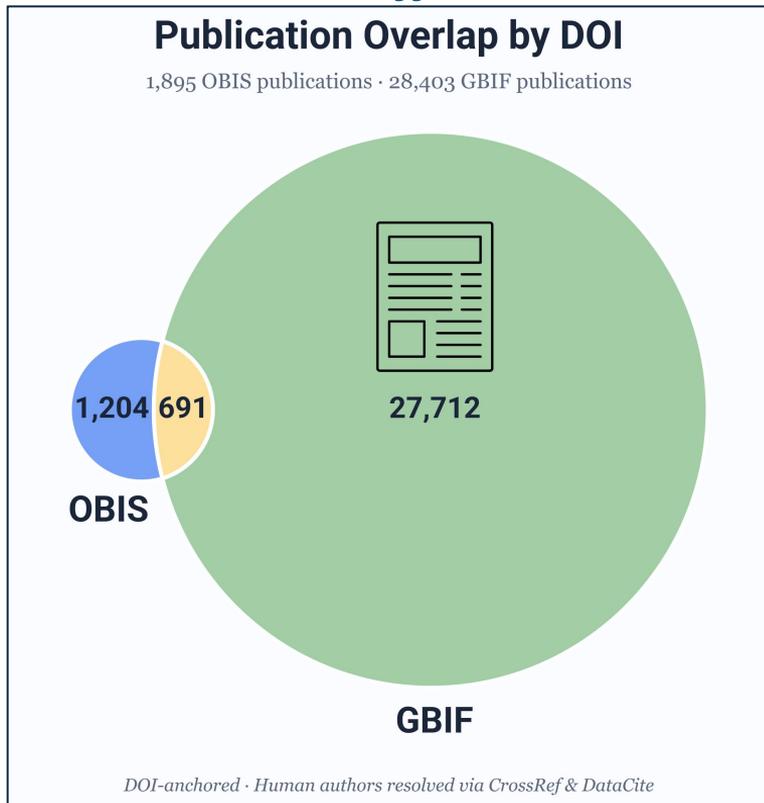
OBIS data supports

crucial global & regional **ocean assessments** to enable science-based decision making



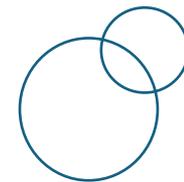
OBIS-GBIF Communities

The same, but different.

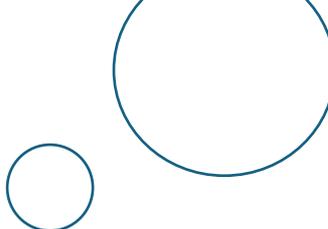




OBIS : GBIF Data Differences



	OBIS	GBIF
Core tables	Occurrence, Event	Occurrence, Event, Taxon
Taxonomic Backbone	WoRMS	Catalogue of Life
Extension tables in downloads	Yes - prompted (eMoF, DNA)	Optional filter or via API
Quality Control differences	Depth, WoRMS IDs, habitat (marine, FW, terrestrial)	Interprets submitted data



DISCUSSION & QUESTIONS



OBIS-GBIF Marine Data Workflow

1

Check OBIS
required fields



2

Ensure taxa
align to
WoRMS

- a. Include aphiaIDs to
scientificNameID if
possible



3

Add OBIS as a
Network on
IPT



Minimum data requirements

OBIS required DwC terms

1. **eventDate**
2. **eventID**
3. **decimalLatitude**
4. **decimalLongitude**
5. **occurrenceID**
6. **occurrenceStatus**
7. **basisOfRecord**
8. **scientificName**
9. **scientificNameID** (strongly recommended)
10. **minimumDepthInMeters** (strongly recommended)
11. **maximumDepthInMeters** (strongly recommended)

Green = shared with GBIF



Credit: Philip Hamilton / Ocean Image Bank

Aligning Taxa to WoRMS



WoRMS Taxon match

You can use the WoRMS Taxon Match Tool ([credits](#)) to automatically match your species list or taxon list with WoRMS. After matching, the tool will return your file with the AphiaID's, valid names, authorities, WoRMS classification and/or any other output you selected. [\[View manual\]](#)
For performance reasons, the limit is set to 1,500 rows. For matching larger files, non-marine or multiple datasources, please use the [Lifewatch Species Information Backbone](#).

File: worms_tax...tch (test) csv

Allowed filetypes: Plain text [TXT], Comma Separated [CSV] & Excel sheet [XLS, XLSX]

Row delimiter: First row contains column names

Column delimiter:

Match authority:

Match upto: Higher taxa only possible if a full classification is given in additional columns

Limit to:
 taxa:
 belonging to:

Output: AphiaID LSID TSN ScientificName Authority Accepted name Classification

- **WoRMS Taxon Match tool**
 - Upload txt or csv of taxa names & receive matches
 - Check LSID box
 - Confirm ambiguous matches
 - Download LSIDs → scientificNameID

WoRMS Taxon match

Match preview for the file 'worms_taxon_match.xlsx' - matching: 90.91% [\[new match\]](#)
If available, please select the [WoRMS](#) taxon that corresponds to your taxon. Then click 'Download'.

ScientificName	ScientificNameAuthor	WoRMS match
Phoronis mulleri	Selys-Longchamps	Phoronis muelleri Selys-Longchamps, 1903
Atylus swammerdami		Atylus swammerdami (H. Milne Edwards, 1830) accepted as Nototropis
Ocenebra erinacea		Ocenebra erinacea (Linnaeus, 1858) accepted as Ocenebra erinaceus (Linnaeus, 1758)
Corbula crassa		<input type="text" value="(ambiguous - select below)"/>
Spio	O.F.Müller, 1776	<input type="text" value="(ambiguous - select below)"/>
Acasta perforata		<input type="text" value="Corbula crassa Reeve, 1843 accepted as Corbula ovalina Lamarck, 1818 [exact]"/>
Acanthias linei		<input type="text" value="Corbula crassa Andree, 1860 [exact]"/>
Anchoviella guianensis	Eigenman, 1912	Anchoviella guianensis (Eigenmann, 1912)
Acanthias linei	Malm, 1877	Acanthias linnei Malm, 1877 accepted as Squalus acanthias Linnaeus, 1758
Mediorhynchus cambellensis		(none)
Ambystoma altamirani		Ambystoma altamirani Dugès, 189

Excel sheet (XLS) Excel sheet (XLSX) Text file SGML

Aligning Taxa to WoRMS

- R or python packages:
 - obistools::match_taxa()
 - worrms::wm_records_taxamatch()
 - Pyworms
- Match taxa, obtain LSIDs
- Merge IDs back into dataset

```
> spp<-read.csv("WoRMS taxon match (test).csv")
> worms<-match_taxa(spp$ScientificName, ask=T)
```

```
9 names, 0 without matches, 1 with multiple matches
```

```
Proceed to resolve names (y/n/info)? y
```

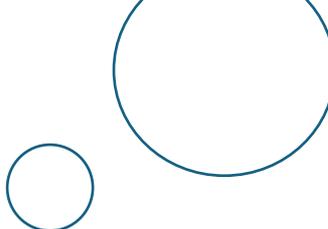
```
# A tibble: 2 × 5
```

	AphiaID	scientificname	authority	status	match_
	<int>	<chr>	<chr>	<chr>	<chr>
1	129625	Spio	Fabricius, 1785	accepted	exact
2	348051	Spio	Brady & Robertson, 1876	unaccepted	exact

```
Spio
```

```
Multiple matches, pick a number or leave empty to skip:
```

scientificName	scientificNameID	match_type	acceptedNameUsageID
Phoronis muelleri	urn:lsid:marinespecies.org:taxname:128549	phonetic	128549
Atylus swammerdamei	urn:lsid:marinespecies.org:taxname:102131	exact	488966
Ocinebra erinacea	urn:lsid:marinespecies.org:taxname:1864778	phonetic	140405
Corbula crassa	urn:lsid:marinespecies.org:taxname:752127	exact	505866
Spio	urn:lsid:marinespecies.org:taxname:129625	exact	129625
Acasta perforata	urn:lsid:marinespecies.org:taxname:718698	exact	718689
Acanthias linnei	urn:lsid:marinespecies.org:taxname:307256	phonetic	105923
Anchoviella guianensis	urn:lsid:marinespecies.org:taxname:275543	phonetic	275543
Acanthias linnei	urn:lsid:marinespecies.org:taxname:307256	phonetic	105923



DISCUSSION & QUESTIONS



Add GBIF Data to OBIS

1. Ensure dataset is **registered** w/GBIF
2. Add OBIS as a **Network**
3. GitHub issue created @ github.com/iobis/obis-network-datasets
4. OBIS Node will **endorse** and dataset added to OBIS

❑ **Not applicable to IPTs already indexed by OBIS**

Registration

Register

GBIF registration data.



Harmful Algal Event Database (HAEDAT)

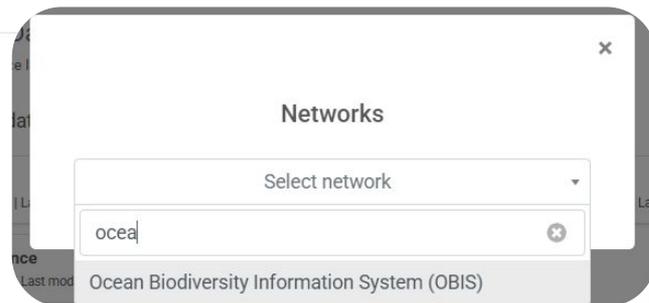
90dbe62a-5470-4aa9-b285-059e95b49936 | IOC Harmful Algal Bloom Programme



Networks

+ Add

No GBIF networks associated with the resource.



Add GBIF Data to OBIS

<https://www.youtube.com/watch?v=4U1mjvCpC6s&feature=youtu.be>





Summary



OBIS-GBIF Publishing

- Reach wider communities
 - Support ocean assessments
 - Increase visibility
-
- Align w/required terms, taxa to WoRMS
 - Add GBIF datasets to OBIS network

Engage with OBIS

Work with an OBIS Node or become a
GBIF-OBIS Node

helpdesk@obis.org

Explore our data and our community

obis.org
helpdesk@obis.org
manual.obis.org





Resource List

OBIS Data & Publishing Resources

- <https://manual.obis.org/>
- <https://www.youtube.com/@oceanbiodiversityinformati6931>
- wktmap.com
- maptool.obis.org
- obis.shinyapps.io/coordinates
- github.com/iobis/obis-network-datasets
- <https://github.com/iobis/robis>
- <https://github.com/iobis/obistools>
- <https://rshiny.lifewatch.be/BioCheck/>
- https://ioos.github.io/bio_data_guide/
- <https://sequence.obis.org/>
- <https://shiny.obis.org/distmaps/>

Vocabulary

- https://vocab.nerc.ac.uk/search_nvs/
- <https://vocab.seadatanet.org/p01-facet-search>
- <https://github.com/nvs-vocabs/OBISVocabs/issues>
- <https://mof.obis.org/>

Metadata

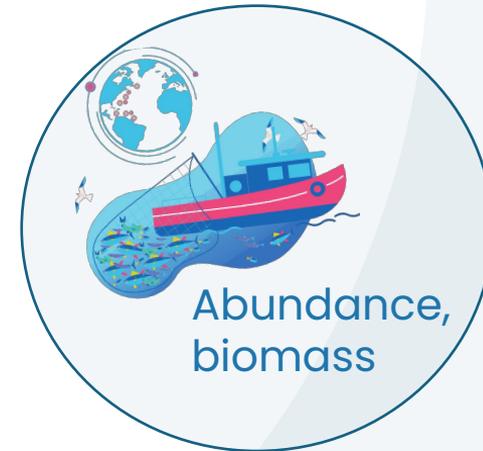
- <https://bioeco.goosocean.org/>
- <https://eovmetadata.obis.org/>



OBIS

More than just species occurrences

>358,000,000 measurements



Data Type

- 📍 location ●
- 🐟 species occurrence ●
- 👥 socio-economic ●
- 🌡️ measurements or variables ●
- 🌿 habitat ● ●
- 🚤 sampling information ● ●
- 📹 imaging, video, acoustic ● ● ●
- 🧬 DNA ● ● ● ●

Data Table

- Event
- Occurrence
- eMoF
- DNA

extended Measurement Or Fact (eMoF) table

- Extension table to record all types of measurements (abiotic, biotic, sampling...)



eDNA sampling
<https://unesdoc.unesco.org/ark:/48223/pf0000384014> (C)
UNESCO/Raw Visuals LTD and (C) UNESCO/Ward Appeltans



Credit: Shaun Wolfe / Ocean
Image Bank

extendedMeasurementOrFact (eMoF) table

- Extension table to record all types of measurements (abiotic, biotic, sampling...)
- Measurements in **long format**

occurrenceID	Length (um)	Temperature (C)	Net
spp1	15	10	30cm Denmark AS small plankton net
spp2	30	12	30cm Denmark AS small plankton net
spp3	23	11	100 cm Denmark AS small plankton net



occurrenceID	measurementType	measurementValue	measurementUnit
spp1	length	15	um
spp1	net diameter	30	cm
spp1	net length	60	cm
spp1	temperature	10	C
spp1	name of net	KC Denmark AS small plankton net	



extendedMeasurementOrFact (eMoF) table

- Extension table to record all types of measurements (abiotic, biotic, sampling...)
- Measurements in **long format**
- Links to **Event** and **Occurrence tables**
- measurementType recorded as *free text* = **heterogeneity**

len	Inth		carapace length	fish length	
	length	standard length	longueur		largo
mean length		fork length	survey length	total length	

eMoF Controlled Vocabulary

- Identifiers for **each** eMoF column (as applicable)
 - measurementUnitID
 - measurementValueID
 - non-numeric values
 - measurementTypeID
- Facilitates data **understanding**
- Enables **data aggregation**
- **Decreases misuse potential**

measurementType	measurementTypeID
length	http://vocab.nerc.ac.uk/collection/P01/current/OBSINDLX/
net diameter	http://vocab.nerc.ac.uk/collection/P01/current/DSAMPA01/
measurementUnit	measurementUnitID
cm	http://vocab.nerc.ac.uk/collection/P06/current/ULCM/
Celsius	http://vocab.nerc.ac.uk/collection/P06/current/UPAA/

measurementValue	measurementValueID
KC Denmark AS small plankton net	http://vocab.nerc.ac.uk/collection/L22/current/NETT01

eMoF Controlled Vocabulary

The NERC Vocabulary Server (NVS)

Service Status

- Identifiers for **each** eMoF column (as applicable)
 - measurementUnitID
 - measurementValueID
 - non-numeric values
 - measurementTypeID
- **OBIS** recommends populating these with **NERC vocabulary collections**

NVS Home | Vocabularies | Thesauri | Search NVS | SPARQL | Other Tools | About NVS

Search for a term in a vocabulary collection

Enter search string using % as wildcard if required. Example: chlorophyll%sediment.

Identifier Preferred label Alternative label Definition Exact match Case sensitive [toggle advanced options](#)

A01	A02	A03	A04	A05	B02	B03	B04	B05	B06	B07	B09	B11	B12	B20	B21	B22	B39	B75	B76	BQ1	BQ2	BQ3	C00	C10	C16
C17	C18	C19	C22	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39	C40	C41	C43	C45	C46	C47	C48	C59	C60	C61	C62	C64
C67	C71	C72	C75	C77	C86	C87	C88	C89	C96	C98	D01	E01	E02	EF1	EL1	EL2	F02	G01	G02	G03	G04	G05	G06	G07	G08
G09	G10	G11	G12	G13	G14	G15	G17	G18	G20	G21	G22	G23	G25	G26	G28	G29	G30	GBX	GBB	GBS	GGT	GS1	GS2	GS3	GS4
GS5	GS6	GS8	GS9	GSA	GSB	GSC	GXM	H01	H02	H03	H04	H05	H06	HA2	I01	I02	I03	I10	I11	I12	I13	I14	I15	L02	L03
L04	L05	L06	L07	L08	L10	L11	L12	L13	L14	L15	L18	L19	L20	L21	L22	L23	L24	L26	L27	L30	L31	L33	L34	L35	M01
M03	M04	M05	M06	M09	M10	M11	M12	M13	M14	M15	M16	M17	M18	M20	M21	M22	M23	M24	M25	M26	M27	M28	M29	MVB	N01
N02	N03	N04	N05	N06	OD1	OG1	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10	P11	P12	P13	P14	P15	P17	P18	P19	P20
P21	P22	P23	P24	P25	P26	P27	P28	P29	P30	P31	P33	P35	P36	P37	P38	P64	Q01	R01	R03	R04	R05	R06	R07	R08	R09
R10	R11	R12	R13	R14	R15	R16	R18	R19	R20	R21	R22	R23	R24	R25	R26	R27	R28	R40	RD2	RMC	RP2	RR2	RTV	S01	S02
S03	S04	S05	S06	S07	S09	S10	S11	S12	S13	S14	S15	S18	S19	S20	S21	S22	S23	S24	S25	S26	S27	S28	S29	S30	T01
V12	V22	V23	W01	W02	W03	W04	W05	W06	W07	W08	W09	W10	W11												

Vocabulary collection selector: hover on the coloured cells to see the collection's title and click to select. Note that the codes and the colours have no meaning but related vocabularies tend to be given a code starting with the same letter.

Search for a term across vocabulary collections

Enter search string

Identifier Preferred label Alternative label Definition Exact match Case sensitive

Search for vocabulary collections

Enter search string using % as wildcard if required. Example: parameter%vocabulary.

Identifier Title Short title Description Governance Exact match Case sensitive

https://vocab.nerc.ac.uk/search_nv

eMoF Controlled Vocabulary

- Identifiers for **each** eMoF column (as applicable)
 - measurementUnitID
 - measurementValueID
 - non-numeric values
 - measurementTypeID
- **OBIS recommends populating these with NERC vocabulary collections**

4	L15	L18	L19	L20	L21	L22
4	M15	M16	M17	M18	M20	M21
3	P04	P05	P06	P07	P08	P09
0	P31	P33	P35	P36	P37	P38
0	R21	R22	R23	R24	R25	R26

eMoF Controlled Vocabulary

- Identifiers for **each** eMoF column (as applicable)
 - measurementUnitID
 - measurementValueID
 - non-numeric values
 - measurementTypeID
- **OBIS recommends populating these with NERC vocabulary collections**

31	C32	C33	C34	C35	C36	C37	C38	C39	C40	C41
86	C87	C88	C89	C96	C98	D01	E01	E02	EF1	EL1
14	G15	G17	G18	G20	G21	G22	G23	G25	G26	G28
5B	GSC	GXM	H01	H02	H03	H04	H05	H06	HA2	I01
10	L11	L12	L13	L14	L15	L18	L19	L20	L21	L22
10	M11	M12	M13	M14	M15	M16	M17	M18	M20	M21
01	OG1	P01	P02	P03	P04	P05	P06	P07	P08	P09
26	P27	P28	P29	P30	P31	P33	P35	P36	P37	P38
15	R16	R18	R19	R20	R21	R22	R23	R24	R25	R26
09	S10	S11	S12	S13	S14	S15	S18	S19	S20	S21
02	W04	W05	W06	W07	W08	W09	W10	W11	W12	W13

Many collections, e.g. C35, L22, L05, C
S11... https://vocab.nerc.ac.uk/search_nv

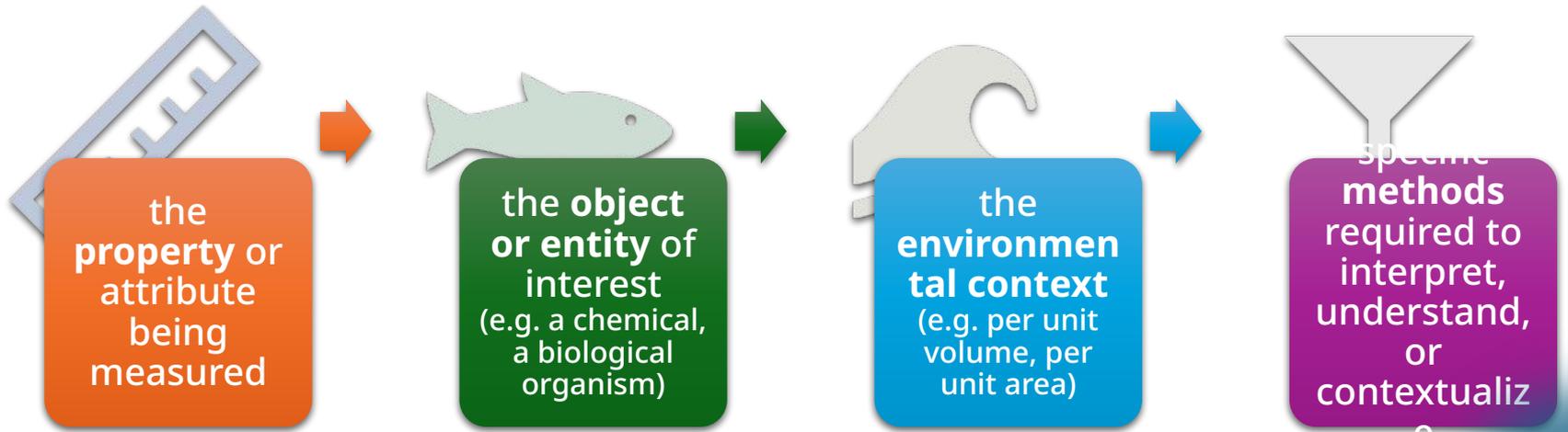
eMoF Controlled Vocabulary

- Identifiers for **each** eMoF column (as applicable)
 - measurementUnitID
 - measurementValueID
 - non-numeric values
 - measurementTypeID
- **OBIS recommends populating these with NERC vocabulary collections**

08	L10	L11	L12	L13	L14	L15
09	M10	M11	M12	M13	M14	M15
00	Q01	Q02	P01	P02	P03	P04
25	P26	P27	P28	P29	P30	P31
14	R15	R16	R18	R19	R20	R21

What info should vocabulary terms include?

- Vocab labels are composed of different elements
 - Not all are mandatory



How to find & choose vocabulary codes

- measurementUnitID and measurementValueID
 - Their **respective NERC vocabulary collections**
- measurementTypeID
 - **STRONGLY recommend SeaDataNet P01 facet search**



<https://vocab.seadatanet.org/p01-facet-search>

The NERC Vocabulary Server (NVS) Service Status

NVS Home | Vocabularies | Thesauri | Search NVS | SPARQL | Other Tools | About NVS

Search for a term in a vocabulary collection

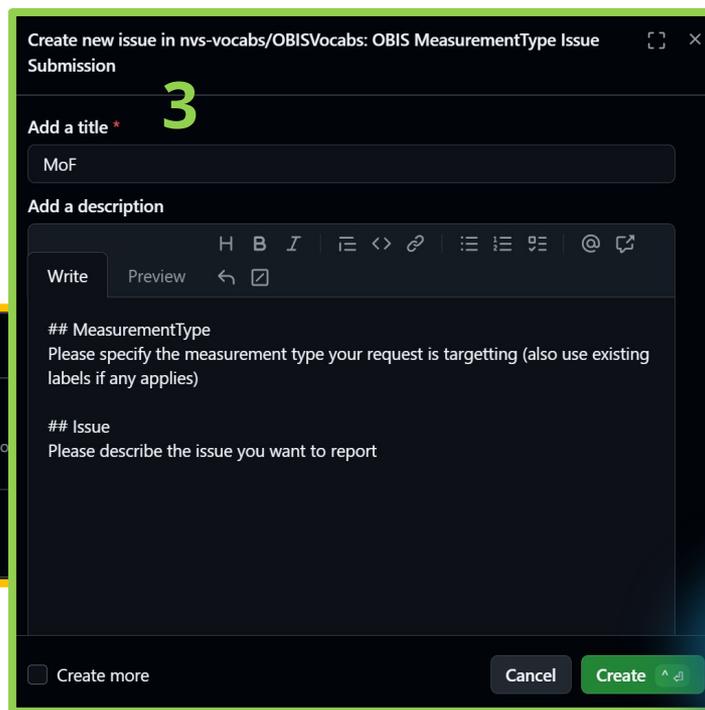
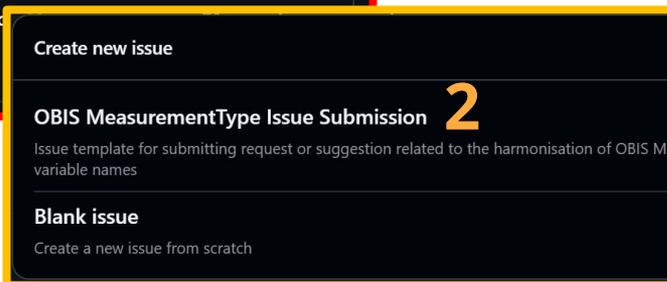
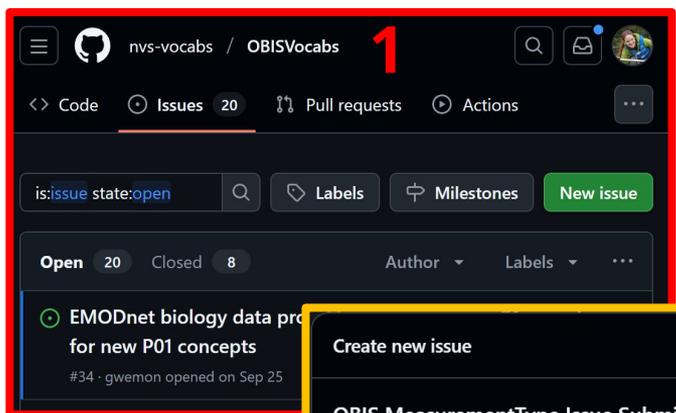
Enter search string using % as wildcard if required. Example: chlorophyll%sediment. Vocab ID Search

Identifier Preferred label Alternative label Definition Exact match Case sensitive [toggle advanced options](#)

A01	A02	A03	A04	A05	B02	B03	B04	B05	B06	B07	B09	B11	B12	B20	B21	B22	B39	B75	B76	BQ1	BQ2	BQ3	C00	C10	C16
C17	C18	C19	C22	C30	C31	C32	C33	C34	C35	C36	C37	C38	C39	C40	C41	C43	C45	C46	C47	C48	C59	C60	C61	C62	C64
C67	C71	C72	C75	C77	C86	C87	C88	C89	C96	C98	D01	E01	E02	EF1	EL1	EL2	F02	G01	G02	G03	G04	G05	G06	G07	G08
G09	G10	G11	G12	G13	G14	G15	G17	G18	G20	G21	G22	G23	G25	G26	G28	G29	G30	GBX	GGB	GGG	GGT	GS1	GS2	GS3	GS4
GS5	GS6	GS8	GS9	GSA	GSB	GSC	GXM	H01	H02	H03	H04	H05	H06	HA2	I01	I02	I03	I10	I11	I12	I13	I14	I15	L02	L03
L04	L05	L06	L07	L08	L10	L11	L12	L13	L14	L15	L18	L19	L20	L21	L22	L23	L24	L26	L27	L30	L31	L33	L34	L35	M01

Request new code

- Submit GitHub ticket: <https://github.com/nvs-vocabs/OBISVocabs/issues>



Measurements in OBIS

Measurement types

Search

measurementType	measurementTypeID	prefLabel	records ↓
Sampling instrument name	http://vocab.nerc.ac.uk/collection/Q01/current/Q0100002/ Q01	<input type="text" value="find"/>	16,835,785
visibility			8,893,974
temperature			8,210,122
fish length			7,992,276
underwater visibility			7,992,276
Abundance	http://vocab.nerc.ac.uk/collection/P01/current/SDBIOL01/ P01	<input type="text" value="find"/>	7,851,455
surge			7,667,787
Number of species observed during time period			7,268,666



CAUTION: NOT VALIDATED

<https://mof.obis.org/>

Manuals

Tools

Vocabulary

OBIS Manual

manual.obis.org

- Comprehensive guidelines
- DwC Term checklist
- Example datasets
- OBIS & GBIF guidelines

How-To Videos



- 18 videos on data formatting and publishing
 - 7 translated to Spanish

<https://www.youtube.com/@oceanbiodiversityinformati6931>

Manuals

Tools

Vocabulary

Coding & QC Tools

- R & Python packages
 - robis, obistools
- Data QC Tools
 - EMODNet Biocheck Tool
- OBIS notebooks
 - Parquet data exports, speciesgrids
- Community-created code examples
 - E.g. IOOS BioData Guide



Georeference tools

- Map and geometry string tools
 - wktmap.com
 - maptool.obis.org
- Shiny app coordinate converters
 - obis.shinyapps.io/coordinates

Metadata

- GOOS BioEco Portal
- EOV Metadata Application

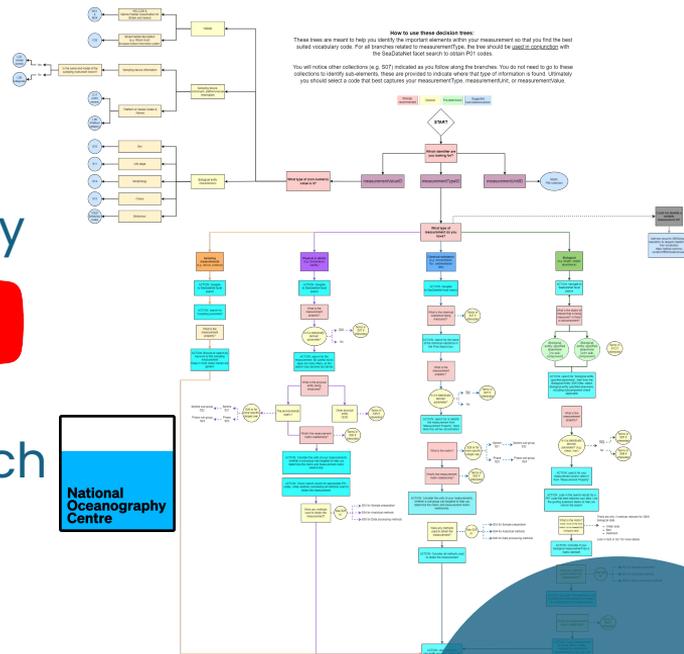
Manuals

Tools

Vocabulary

Controlled vocabulary

- OBIS Vocabulary Team
- Tools for finding vocabulary in measurement data
 - OBIS decision tree
 - How-To YouTube series
 - SeaDataNet Facet Search



OBIS Supports...

Services

- DOIs for download subsets
 - obis.org/doi
- DNA Sequence Search (prototype)
 - sequence.obis.org
- Invasive Species Decision Support tool (PacMAN)
 - portal.pacman.obis.org
- MPA species distribution maps
 - shiny.obis.org/distmaps

Ocean Assessments

OBIS supports crucial global & regional ocean assessments to enable science-based decision making

