

# **OBIS Manual**

# Table of contents

<b>Making better use of OBIS data</b>	<b>4</b>
<b>I Overview</b>	<b>5</b>
<b>1 Introduction</b>	<b>6</b>
1.1 Guidelines on the sharing and use of data in OBIS . . . . .	6
1.2 Acknowledgements . . . . .	6
1.3 Data Policy . . . . .	6
1.3.1 Guidelines on the sharing and use of data in OBIS . . . . .	6
1.4 Getting Help in OBIS . . . . .	8
<b>II Contributing data to OBIS</b>	<b>10</b>
<b>2 What can you contribute and how?</b>	<b>11</b>
2.1 Why publish data to OBIS . . . . .	12
2.2 How to handle sensitive data . . . . .	13
<b>III Data Formatting</b>	<b>15</b>
<b>3 Dataset structure</b>	<b>16</b>
3.1 When to use Event Core . . . . .	19
3.2 When to use Occurrence Core . . . . .	19
3.3 Extensions in OBIS . . . . .	20
3.3.1 How are extensions linked to core tables in OBIS? . . . . .	20
3.3.2 Differences between identifiers . . . . .	20
3.4 Data formatting tools . . . . .	21
<b>4 Formatting data tables</b>	<b>22</b>
4.1 Darwin Core Term Checklist for OBIS . . . . .	22

<b>IV Ensuring Data Quality</b>	<b>30</b>
<b>5 Data quality control</b>	<b>31</b>
5.1 Why are records dropped? . . . . .	31
5.2 How to conduct Quality Control . . . . .	32
5.2.1 Conducting QC with obistools . . . . .	33
5.2.2 QC with R package Hmisc . . . . .	34
<b>V Publishing Data</b>	<b>35</b>
<b>6 Data publication and sharing</b>	<b>36</b>
6.1 Licenses . . . . .	37
<b>VI Access Data from OBIS</b>	<b>38</b>
<b>7 Data access</b>	<b>39</b>
7.1 OBIS Homepage and dataset pages . . . . .	39
7.2 Mapper . . . . .	41
7.3 R package . . . . .	42
7.4 API . . . . .	43
7.5 Full exports . . . . .	43
7.6 Finding your own data in OBIS . . . . .	44
7.7 How to contact data provider . . . . .	45
7.8 Interpreting downloaded files from OBIS . . . . .	45
<b>VII Data Visualization and Analysis</b>	<b>47</b>
<b>8 Data Visualization</b>	<b>48</b>
8.1 Example notebooks using data from OBIS . . . . .	48
8.2 obisindicators: calculating & visualizing spatial biodiversity using data from OBIS . . . . .	48
<b>VIII Additional Resources</b>	<b>50</b>
<b>9 Other Resources</b>	<b>51</b>
9.1 MBON Pole to Pole Tutorial . . . . .	51
9.2 IOOS Darwin Core Guide . . . . .	51
9.3 EMODnet Biology . . . . .	51
9.4 Template Generators . . . . .	52

# Making better use of OBIS data

OBIS is not just a repository for biodiversity data. It's also a community of practice that want to leverage the use of ocean biodiversity data. Here you will find tutorials and tools to help you make better use of data from OBIS (and other resources).

You can contribute with tutorials following the guidelines available [at our repository](#).

Don't know where to start? We suggest reading the [OBIS manual](#) where you will find the most complete guide about how OBIS works and the different types of data available.

Check also the available [packages and pipelines](#) for R and other languages, and [our tool](#) to find the right Darwin Core schema for your data.

# **Part I**

## **Overview**

# 1 Introduction

This manual provides an overview on how to contribute data to OBIS and how to access data from OBIS. It provides guidelines for OBIS nodes and data providers on the OBIS standards and data management best practices to ensure that data published via OBIS are of high quality and follows internationally recognised standards. It also provides guidelines for data users on how to access, process and visualize data from OBIS.

The OBIS manual is a dynamic document and is revised on a regular basis. Suggestions for additions and changes to this document are welcome and can be sent to the OBIS Capacity Development Task Team by email to [training@obis.org](mailto:training@obis.org) or added as issues at <https://github.com/iobis/manual/issues>.

## 1.1 Guidelines on the sharing and use of data in OBIS

It is important that our data providers as well as all the data users are aware and agree on the [OBIS guidelines on the sharing and use of data in OBIS](#), which was adopted at the [4th OBIS Steering Group](#).

## 1.2 Acknowledgements

This manual received contributions from: [Leen Vandepitte](#), [Mary Kennedy](#), [Philip Goldstein](#), [Pieter Provoost](#), [Samuel Bosch](#), [Ward Appeltans](#), [Abby Benson](#), [Yi-Ming Gan](#), [Carolina Peralta Brichtova](#), [Saara Suominen](#), [Serita van der Wal](#), and [Elizabeth Lawrence](#).

## 1.3 Data Policy

### 1.3.1 Guidelines on the sharing and use of data in OBIS

Adopted at SG-OBIS-IV (Feb 2015) and IODE-XXIII (March 2015).

The OBIS data policy is based on the principles of timely, free and unrestricted access to biodiversity data for the benefit of science and society, as defined in the:

- [IOC data exchange policy](#)
- [IOC guidelines on transfer of marine technology](#)
- [IODE objectives](#)
- [OBIS vision and mission](#)

Unless data are collected through activities funded by IOC/IODE, neither UNESCO, IOC, IODE, the OBIS Secretariat, nor its employees or contractors, own the data in OBIS and they take no responsibility for the quality of data or products based on OBIS, or the use or misuse that people may make of them nor can it control or limit the use of any data or products accessible through its website, other than through the use of a published Data Sharing and Use Terms and Conditions.

#### **1.3.1.1 Data sharing agreement**

The data providers retain all rights and responsibilities associated with the data they make available to OBIS via the OBIS nodes. The OBIS nodes warrant that they have made the necessary agreements with the original data providers that it can make the data available to OBIS data under the [Creative Commons licenses](#).

The data providers are responsible for the completeness of the data and metadata profiles. When data is made available to OBIS, OBIS is granted permission to:

- Distribute the data via its data and information portal
- Build an integrated database, use the data for data quality control purposes, complement the data with other data such as climate variables and build value-added information products and services for science and decision-making
- Serve the data to other similar open-access networks such as GBIF in compliance with the terms and conditions for use set by the data providers.

In pursuance of copyright compliance, OBIS endeavours to secure permission from rights holders to ingest their datasets. In the event that the inclusion of a dataset in OBIS is challenged on the basis of copyright infringement, OBIS will follow a take-down policy until there is resolution.

#### **1.3.1.2 Data use agreement**

The data in OBIS are freely available to everyone, following the principles of equitable access and benefit sharing and supporting capacity development and participation of all IOC Member States in global programmes. However, data users are expected to give attribution to the data providers (see Citations) and the use of data from OBIS should happen in the light of fair use, i.e.:

- Recognize that the OBIS portal holds the master copy of the integrated database and hence users should refrain from online redistribution of the OBIS database. Because the OBIS database is updated regularly (every so months) with new datasets and revisions of existing datasets, copies of the OBIS database will become out of date quickly. If you wish to build access web services on top of OBIS, please contact the [OBIS secretariat](#).
- Respect the data providers, and provide helpful feedback on data quality.
- In the case you are a custodian of biogeographic data yourself you should take action to also publish these data through OBIS.
- Consider sponsoring or partnering with OBIS and its OBIS nodes in grant proposal writing. Creating a global database like OBIS cannot happen without the, often voluntary, contribution of many scientists and data managers all over the world. Several activities, such as the coordination, data aggregation, quality control, database and website maintenance require resources including manpower at national and international level. A list of sponsors can be found [here](#)

### 1.3.1.3 Disclaimer

Appropriate caution is necessary in the interpretation of results derived from OBIS. Users must recognize that the analysis and interpretation of data require background knowledge and expertise about marine biodiversity (including ecosystems and taxonomy). Users should be aware of possible errors, including in the use of species names, geo-referencing, data handling, and mapping. They should crosscheck their results for possible errors, and qualify their interpretation of any results accordingly.

Unless data are collected through activities funded by IOC/IODE, neither UNESCO, IOC, IODE, the OBIS Secretariat, nor its employees or contractors, own the data in OBIS and they take no responsibility for the quality of data or products based on OBIS, or the use or misuse.

## 1.4 Getting Help in OBIS

If you require additional assistance with OBIS we recommend you first get in touch with the most [relevant OBIS node](#). We also have a **support channel** on [Slack](#) where you can communicate with the OBIS community for help. Please feel comfortable posting to this channel before reaching out to the OBIS Secretariat ([helpdesk@obis.org](mailto:helpdesk@obis.org)). The OBIS community is quite active on Slack and GitHub (see below) so you are more likely to receive a quick answer to your question by posting in either place, as the Secretariat receives many requests.

You can submit issues and questions on relevant Github repositories:

- [OBIS Manual](#)
- [OBIS issues GitHub repo](#)



- [OBIS quality control issues](#)
- [All other OBIS repositories](#)

We strongly recommend creating a GitHub account to engage with the OBIS community, document issues, ask questions, find datasets that need endorsing, etc. GitHub gives threads a more permanent home and allows for open communication and transparency. If you are unfamiliar with GitHub, the Carpentries have [these training resources](#) which you can reference.

## **Part II**

# **Contributing data to OBIS**

## 2 What can you contribute and how?

Since 2000, OBIS has accepted, curated and published marine biodiversity data obtained by varied sources and methods. There is a common misconception that OBIS only accepts species occurrence data - however this is not true! OBIS can accept many types of marine data including:

- Presence/Absence
- Abundance, individual count
- Biomass
- Abiotic measurements
- Biotic measurements
- Sampling methods
- Sample processing methods
- Genetic data including sequences
- Data originating from [historical records](#)
- Tracking data
- Habitat data
- Acoustic data
- Imaging data
- Metadata describing the dataset and any project or programme related metadata

So if you have any of these types of marine data linked to your occurrence data and also want to contribute to OBIS - great! OBIS accepts data from any organization, consortium, project or individual who wants to contribute data. OBIS Data Sources are the authors, editors, and/or organisations that have published one or more datasets through OBIS. They remain the owners or custodians of the data, not OBIS!

OBIS harvests and publishes data from recognized IPTs from OBIS nodes or GBIF publishers. If you own data or have the right to publish data in OBIS, you can contact the [OBIS secretariat or one of the OBIS nodes](#), or additionally a GBIF publisher. Your organization or programme can also [become an OBIS node](#). An OBIS node usually publishes data from multiple data holders, effectively being a node in a network of data providers. So you may have to first find a [relevant node](#) before you get your data ready to publish.

To publish a dataset to OBIS, there are **five** main steps you must go through.

1. First, you must [identify](#) which OBIS node is best suited to host your published data. If you would like to [publish to GBIF](#) at the same time, that is also possible. If your

organization is already affiliated with a GBIF node with which you must publish from, OBIS can also [harvest from GBIF nodes](#).

2. Second, you must determine the [structure](#) of your data and which format will best suit your dataset. OBIS follows Darwin Core Archive (DwC-A) standards for datasets, and currently follows a star schema format. This format is based on relational databases. If you are unfamiliar with such database structures, or would like to refamiliarize yourself with them, please read [here](#)
3. Then, you need to actually [format](#) your data according to OBIS and DwC-A standards and guidelines
4. Once formatted, you should run a series of [quality control](#) measures to ensure you are not missing any required information and that all standards are being met. This helps ensure all data published in OBIS is formatted in a standardized way. When published in OBIS, OBIS provides a quality report to inform data owners and users of any quality control issues. By completing quality control before you publish your dataset you ensure there are fewer errors to fix later.
5. Now that your dataset is ready for publishing, the relevant metadata must be filled in, and then published on the previously identified IPT.

Each of these steps are covered in detail in the relevant sections of the manual. For an overview of this process see [data management flow in OBIS](#).

## 2.1 Why publish data to OBIS

It is important to publish and ensure your dataset follows a universal standard for several reasons. The [FAIR guiding principles](#) for scientific data management and stewardship provide a good framework to understand the reasoning behind publishing data. FAIR stands for Findable, Accessible, Interoperable, and Reusable. Let's understand each aspect within the FAIR framework and how it is linked to publishing data in OBIS.

- **F - Findable**

Even if you publish your dataset on its own, publishing your data with OBIS will make your data more Findable (and Accessible) to a wider audience you might not have otherwise reached. By publishing your dataset to OBIS you are adding to a global database where your data can be found and analyzed alongside thousands of other datasets. For example, a dataset on [marine invasive species in Venezuela](#) was published July 20, 2022 and as of October 5, 2022 records of this dataset were included in 1,873 data download requests. This can save you time rather than handling individual data requests.

- **A - Accessible**

Similar to being Findable, OBIS makes your datasets more Accessible. Each dataset is given an identifier when you upload it on an IPT. Thus when users obtain data from OBIS, the original dataset can easily be identified and accessed. Data from OBIS is accessible in [numerous ways](#), giving data users multiple avenues to potentially access your data.

- **I - Interoperable**

Using a standardized data format with controlled vocabularies will ensure your data are more Interoperable - more easily interpreted and processed by computers and humans alike. Increasingly, scientists use computer programs to conduct e-Science and collect data with algorithms. Formatting your data for OBIS will ensure it can be read and accessed by such programs as well as understood by users.

- **R - Reusable**

Publishing your data allows it to be Reused according to your chosen [data usage license](#). Very likely you expended resources to collect your data and it would be a waste of those resources to leave your unique data unpublished and inaccessible for current and future generations. Likewise, it is better to preserve any data processing done to ensure your dataset is reproducible and/or verifiable. Finally, data in OBIS is often used in several assessment processes and used as information to support policy makers around the globe making informed decisions.

There are many other benefits of publishing in OBIS, even if you haven't published any work on it yet. This includes:

- Your dataset can be [associated with a DOI](#), allowing for your dataset to be more easily cited. By ensuring your dataset citation is complete you will ensure you are being cited properly.
- Publishing your dataset with OBIS makes it easier to set it up as a [Data paper](#), which generates value for you and other researchers.
- There are social benefits to data publishing as your work becomes integrated into a wider dataset. It gives both you and your data more visibility. This can lead to more opportunities for collaboration and further career development as a researcher or professional.
- Your data can be incorporated into larger analyses to better understand global ocean biodiversity, helping to shape regional and international policies.

## 2.2 How to handle sensitive data

We recognize that sometimes your dataset may contain sensitive information (e.g., location data on endangered or poached species), or perhaps your organization does not want certain details publicly accessible. Types of sensitive data include:

- Location data on endangered or protected species

- Information regarding a commonly poached species
- Species or locations that have an economic impact (positive or negative)

To accommodate sensitivity but still be able to contribute to OBIS, we suggest:

- [Generalizing location](#) information by: Obtaining regional coordinates using [MarineRegions](#), [Getty Thesaurus of Geographic Names](#), or [Google Maps](#)
- Using the [OBIS Map tool](#) to generate a polygon area with a Well-Known Text (WKT) representation of the geometry to paste into the `footprintWKT` field.
- Delay timing of publication (e.g., to accommodate mobile species)
- [Submit your dataset, but mark it as private in the IPT](#) so it is not published right away (i.e., until you set it as public). Alternatively, you can set a password on your dataset in order to share with specific individuals. Note that setting passwords will require some coordination with the IPT manager. By submitting your data to an IPT but not immediately publishing it, you can ensure that the dataset will be in a place to be incorporated at a later date when it is ready to be made public. This not only saves time and helps retain details while relatively fresh in your mind, but also ensures the dataset is still ready to be mobilized in case jobs are changed at a later date.

GBIF has created the following [Best Practices for Generalizing Sensitive data](#) which can provide you with additional guidance. Chapman AD (2020) Current Best Practices for Generalizing Sensitive Species Occurrence Data. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc-5jp4-5g10>.

## **Part III**

# **Data Formatting**

### 3 Dataset structure

Formatting data can be challenging. This section of the manual deals with how to format data for OBIS, beginning with an overview of dataset structure.

Deciding on your dataset structure is one of the first steps towards getting your data ready for publishing. At this step, there are different non arbitrary you need to do with your data, but it is important to determine which structure best suits your dataset before proceeding. Then, once you have decided on the dataset structure, you can continue formatting your data.

We have created the following flow chart for an overview on how to determine what structure best suits your data.

What kind of data do you have, or will collect?

What kind of data do...

occurrence

occurrence

genetic

genetic

biomass

biomass

tracking

tracking

habitat

habitat

Do you have event level information (e.g. sample, station, cruise, study, etc.)?

Do you have event level infor...

abundance, percent cover

abundance, perce...

Yes



Yes

No

No

Have (or will) you collect any data associated with samples or sampling? (e.g. temperature, length, etc.)

Have (or will) you collec...

Event core

Event core

Occurrence Core

Occurrence Core

Yes

Yes

No

No

Event core

Event core

Event core only

Event core...

Occurrence Core

Occurrence Core

Have (or will) you collect any data associated with samples or sampling? (e.g. temperature, length, etc.)

Have (or will) you collec...

Yes

Yes

No

No

Occurrence core

Occurrence...

Occurrence core

Occurrence...

DNADerived dataextension

DNA...

Start

Start

eMoFextension

eMoF...

eMoFextension

eMoF...

Have (or will) you collect any data associated with samples or sampling? (e.g. temperature, length, etc.)

Have (or will) you collec...

Yes

Yes

No

No

Occurrencecore

Occurrence...

Occurrencecore only

Occurrence...

eMoFextension

eMoF...

DNADerived dataextension

DNA...

DNADerived dataextension

DNA...

Text is not SVG - cannot display

For more guidance, see the sections below.

### 3.1 When to use Event Core

Event Core describes **when** and **where** a specific sampling event happened and contains information such as location and date. Event Core is often used to organize your data tables when there are more than one sampling occasion and/or location, and different occurrences linked to each sampling. This organization follows the rationale of most ecological studies and typical marine sampling design. It covers:

- When specific details are known about **how** a biological sample was taken and processed. These details can then be defined in the eMoF Extension with the [Q01 vocabulary](#)
- When the dataset contains abiotic measurements, or other biological measurements which are **related to an entire sample** (not a single specimen). For example a biomass measurement for an entire sample, not each species within the sample

Event Core can be used in combination with the Occurrence and eMoF extensions. The identifier that links Event Core to the extension is the `eventID`. `parentID` can also be used to give information on hierarchical sampling. `occurrenceID` can also be used in datasets with Event Core in order to link information between the Occurrence extension and the eMoF extension.

### 3.2 When to use Occurrence Core

Occurrence Core datasets describe **observations** and **specimen records** and cover instances when:

- **No information** on how the data was sampled or how samples were processed is available
- No abiotic measurements are taken or provided
- You have [eDNA and DNA-derived data](#)
- Biological measurements are made on **individual specimens** (each specimen is a single occurrence record)

Occurrence Core is also often the preferred structure for museum collections, citations of occurrences from literature, and sampling activities.

Datasets formatted in Occurrence Core can use the eMoF Extension for when you have biotic measurements or facts about your specimen. The DNA derived data extension can also be used to link to DNA sequences. The identifier that links Occurrence Core to the extension(s) is the `occurrenceID`.

### 3.3 Extensions in OBIS

Currently OBIS accepts the following extensions:

- Occurrence
- Event
- MeasurementOrFact
- extendedMeasurementOrFact
- DNADerivedData

#### 3.3.1 How are extensions linked to core tables in OBIS?

As established in the [relational database section](#), OBIS relies on datasets being formatted according to a relational database structure. The [ENV-DATA approach](#) that OBIS implements means your dataset will have a Core table and (optionally) Extension tables. As a review, a core file contains information relevant and applicable to each record in the extension(s). An extension file then contains records that link back to a record in the core file with more specific information (e.g., methods, measurements, facts, DNA sequences, etc.).

The extension file(s) accepted by OBIS (eMoF, Occurrence, DNA) are linked to your core tables by the use of identifying ID codes. These codes could be either **eventID** or **occurrenceID**. For details on how to construct these IDs, click [here](#).

#### 3.3.2 Differences between identifiers

If your core file is based on occurrences (e.g., a record of one or more taxa specimens), then any extensions are linked with **occurrenceID**. If your core file is based on events (e.g., a sampling event, cruise, observation, etc.), then the linking identifier is **eventID**. In the Core tables, identifiers are always unique, which means, they do not repeat and each line has a different identifier. On the other hand, multiple records in an extension file can have the same identifier which will link them to the same event or occurrence record (depending on which is the Core). The different linking identifiers are shown in the figure below.

Let us consider a fictional plankton trawl sampling event to demonstrate how identifiers link Core and Extension tables in OBIS. This trawl used two types of nets, occurred in March 2013, and has an eventID **plankton-northsea-2013-03**. Suppose we have information about the types of trawl used and the species abundance from this trawling event. The information (e.g., date) of the sampling event itself would be found in the Event Core, whereas the abundance data and sampling methods would be in the eMoF table. How do we ensure the abundance and sampling method data is properly linked to the correct event? By using the same eventID for each record in the eMoF table, **plankton-northsea-2013-03**, the information is properly linked between the Event Core and the eMoF extension.

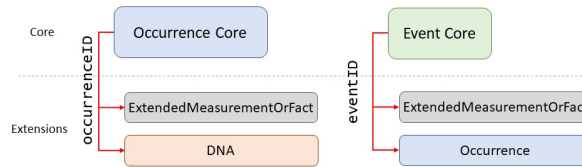


Figure 3.1: Diagram of how the different core tables are linked to their extensions by different identifiers.

### 3.4 Data formatting tools

The GBIF Norwegian Node created the [DwC Excel Template Generator](#). This tool will generate four different types of blank Excel spreadsheets: Occurrence Core, MeasurementOrFact, Metadata, and a README. This tool works best if you already know which Darwin Core fields you need, although a default template can be generated.

Another tool from Norway is the [Excel to Darwin Core Standard \(DwC\) Tool](#). This is a macro Excel spreadsheet that helps create templates for Event (aka Sampling-Event) and Occurrence core tables, as well as MeasurementsOrFacts, Extended MeasurementsOrFacts, and Simple Multimedia extensions. GBIF provides an [Occurrence core template](#) and an [Event core template](#). If you use these templates from GBIF, be aware that [GBIF's required terms are different from OBIS](#).

There are also some tools that can help you unpivot (or flatten) data tables. These can be used to flatten many columns into one, particularly useful for the [eMoF](#) table.

- [GBIF Norway's crosstab to list converter](#). Note that this tool is not completely automated
- [Excel's built-in unpivot function](#)

## 4 Formatting data tables

### 4.1 Darwin Core Term Checklist for OBIS

There are many Darwin Core terms listed in the [TDWG quick reference guide](#). However, not all these terms are necessary for publishing data to OBIS.

For your convenience, we have created a checklist of all the Darwin Core terms relevant for OBIS data providers. You can reference this list to quickly see which terms are required by OBIS, which file (Event, Occurrence, eMoF, DNA) they can be found in, and which Darwin Core class it relates to. These terms correlate with the [IPT vocabulary mapping](#) you will do when it comes time to publish your dataset. You may notice some terms are accepted in multiple data tables (e.g., Event and Occurrence) - this is because it depends on your dataset structure. If you have an Event Core, you will include some terms that would not be included if you had Occurrence Core. For guidance on specific class terms (e.g., location, taxonomy, etc.), see the [Darwin Core](#) section of the manual.

Note that when you publish your dataset on the IPT, if you use a term not listed below it will be an unmapped field and will **not** be published alongside your data. You may still wish to include such fields in your dataset if you are publishing to other repositories, just know that they will not be included in your OBIS dataset. You may include this information either by putting it in the `dynamicProperties` field in JSON format, or putting the information into the `eMoF`. Alternatively, you may have fields that you do not wish to be published and that do not correspond to one of these terms (e.g. personal notes). This is okay - if they are not mapped to one of the terms, that column in your dataset will not be published.

Term	OBIS	DarwinCore			
	Required	Class	Event	Occurrence eMoF	DNA
eventDate	required	event	x	x	
eventID	required	event	x	x	x
decimalLatitude	required	location	x	x	
decimalLongitude	required	location	x	x	
occurrenceID	required	occurrence		x	x
occurrenceStatus	required	occurrence		x	
basisOfRecord	required	record		x	x
scientificName	required	taxon		x	

Term	OBIS Required	DarwinCore Class	Event	Occurrence eMoF	DNA
scientificName	strongly recommended	taxon		x	
DNA_sequence	strongly recommended	dna			x
env_broad_scale	strongly recommended	dna			x
env_local	recommended	dna			x
env_medium	strongly recommended	dna			x
lib_layout	recommended	dna			x
nucl_acid_amp	recommended	dna			x
nucl_acid_ext	recommended	dna			x
otu_class_app	recommended	dna			x
otu_db	recommended	dna			x
otu_seq_comp	recommended	dna			x
pcr_primer_forward	strongly recommended	dna			x
pcr_primer_reverse	strongly recommended	dna			x
pcr_primer_sense	strongly recommended	dna			x
pcr_primer_antisense	strongly recommended	dna			x
samp_name	recommended	dna			x
samp_vol_wet	recommended	dna			x
seq_meth	recommended	dna			x
sop	recommended	dna			x
target_gene	strongly recommended	dna			x
target_subfragment	strongly recommended	dna			x
day	recommended	event	x	x	
endDayOfYear	recommended	event	x	x	
eventRemarks	optional	event	x	x	
eventTime	recommended	event	x	x	
fieldNotes	optional	event	x		
fieldNumber	optional	event	x		

Term	OBIS Required	DarwinCore Class	Event	Occurrence	eMoF	DNA
habitat	recommended	event	x		x	
month	strongly recommended	event	x	x		
parentEventID	Required (if exists)	event	x			
sampleSizeUnit	strongly recommended	event		x	x	
sampleSizeValue	strongly recommended	event		x	x	
samplingEffort	strongly recommended	event		x	x	
samplingProtocol	strongly recommended	event		x	x	
startDayOfYear	recommended	event	x			
verbatimEventDate	recommended	event	x			
year	strongly recommended	event	x	x		
bed	optional	geologicalContext		x		
earliestAgeOrPeriodStage	optional	geologicalContext		x		
earliestEonOrPeriodEonothem	optional	geologicalContext		x		
earliestEpochOrLowestSeries	optional	geologicalContext		x		
earliestEraOrLowestErathem	optional	geologicalContext		x		
earliestPeriodOrLowestSystem	optional	geologicalContext		x		
formation	optional	geologicalContext		x		
group	optional	geologicalContext		x		
highestBiostrophicZone	optional	geologicalContext		x		
latestAgeOrHighestStage	optional	geologicalContext		x		
latestEonOrHighestEonothem	optional	geologicalContext		x		
latestEpochOrHighestSeries	optional	geologicalContext		x		
latestEraOrHighestErathem	optional	geologicalContext		x		
latestPeriodOrHighestSystem	optional	geologicalContext		x		
lithostratigraphicForm	optional	geologicalContext		x		
lowestBiostratigraphicZone	optional	geologicalContext		x		
member	optional	geologicalContext		x		
dateIdentified	optional	identification		x		
identificationID	optional	identification		x		
identificationQualifier	optional	identification		x		



Term	OBIS Required	DarwinCore Class	Event	Occurrence eMoF	DNA
identificationNumber	Optional (required for imaging data)	identification		x	
identificationRemarks	Recommended	identification		x	
identificationVerificationStatus	Optional (required for imaging data)	identification		x	
identifiedBy	optional (required for imaging data)	identification		x	
identifiedByID	Optional	identification		x	
typeStatus	optional	identification		x	
continent	strongly rec- ommended	location	x	x	
coordinatePrecision	strongly rec- ommended	location	x	x	
coordinateUncertaintyInMeters	strongly rec- ommended	location	x	x	
country	recommended	location	x	x	
countryCode	optional	location	x	x	
county	optional	location	x	x	
footprintSpatialFootprint	Optional	location	x	x	
footprintSRID	Optional	location	x	x	
footprintWKID	Recommended	location	x	x	
geodeticDatum	recommended	location	x	x	
georeferencedBy	Optional	location	x	x	
georeferencedDate	Optional	location	x	x	
georeferencePoint	Optional	location	x	x	
georeferenceSpot	Optional	location	x	x	
higherGeography	Optional	location	x	x	
higherGeographyID	Optional	location	x	x	
island	optional	location	x	x	
islandGroup	optional	location	x	x	
locality	recommended	location	x	x	
locationAccordingTo	recommended	location	x	x	
locationID	strongly rec- ommended	location	x	x	

Term	OBIS Required	DarwinCore Class	Event	Occurrence eMoF	DNA
locationRemarks	recommended	location	x	x	
maximumDepthInMeters	recommended	location	x	x	
maximumDistanceAboveSurfaceInMeters	optional	location	x	x	
maximumElevationInMeters	optional	location	x	x	
minimumDepthInMeters	recommended	location	x	x	
minimumDistanceAboveSurfaceInMeters	optional	location	x	x	
minimumElevationInMeters	optional	location	x	x	
municipality	optional	location	x	x	
pointRadiusSpatialUnit	optional	location	x	x	
stateProvince	optional	location	x	x	
verbatimCoordinates	optional	location	x	x	
verbatimCoordinatesSystem	optional	location	x	x	
verbatimDepth	optional	location	x	x	
verbatimElevation	optional	location	x	x	
verbatimLatitude	optional	location	x	x	
verbatimLocality	optional	location	x	x	
verbatimLongitude	optional	location	x	x	
verbatimSR	optional	location	x	x	
waterBody	recommended	location	x	x	
materialSampleID	recommended	materialSample		x	
measurementAccuracy	recommended	measurementOrFact			x
measurementDeterminedBy	optional	measurementOrFact			x
measurementDeterminedDate	optional	measurementOrFact			x
measurementID	recommended	measurementOrFact			x
measurementMethod	recommended	measurementOrFact			x
measurementRemarks	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x
measurementsTypically recommended	recommended	measurementOrFact			x

Term	OBIS Required	DarwinCore Class	Event	Occurrence	eMoF	DNA
associatedMedia	recommended	occurrence		x		
associatedReferences	optional	occurrence		x		
associatedSequences	recommended	occurrence		x		
associatedTaxa	optional	occurrence		x		
behavior	recommended	occurrence		x	x	
catalogNumber	recommended	occurrence		x		
disposition	optional	occurrence		x		
establishmentMeans	optional	occurrence		x		
georeferenceVerificationStatus	optional	occurrence		x		
individualCount	strongly recommended	occurrence		x	x	
lifeStage	recommended	occurrence		x	x	
occurrenceRemarks	recommended	occurrence		x		
organismQuantity	strongly recommended	occurrence		x	x	
organismQuantityType	strongly recommended	occurrence		x	x	
otherCatalogNumbers	optional	occurrence		x		
preparations	optional	occurrence		x		
recordedBy	recommended	occurrence		x		
recordedByID	recommended	occurrence		x		
recordNumber	recommended	occurrence		x		
reproductiveCondition	recommended	occurrence		x		
sex	recommended	occurrence		x	x	
associatedOrganisms	optional	organsim		x		
associatedOrganisms	optional	organsim		x		
organismID	recommended	organsim		x		
organismName	recommended	organsim		x		
organismRemarks	recommended	organsim		x		
organismScope	optional	organsim		x		
previousIdentifications	optional	organsim		x		
accessRights	recommended	record	x	x		
bibliographicCitation	recommended	record	x	x		
collectionCode	optional	record	x	x		
collectionID	optional	record	x	x		
dataGeneralization	optional	record	x	x		
datasetID	recommended	record	x	x		
datasetName	recommended	record	x	x		
dynamicProperties	recommended	record	x	x		
informationWithhold	optional	record	x	x		

Term	OBIS Required	DarwinCore Class	Event	Occurrence eMoF	DNA
institutionCountry	optional	record	x	x	
institutionID	optional	record	x	x	
language	recommended	record	x	x	
license	recommended	record	x	x	
modified	recommended	record	x	x	
ownerInstitutionCode	optional	record	x	x	
references	recommended	record	x	x	
rightsHolder	recommended	record	x	x	
type	strongly recommended	record	x	x	x
acceptedNameUsage	recommended	taxon		x	
acceptedNameUsageID	recommended	taxon		x	
higherClassification	recommended	taxon		x	
infraspecificEpithet	optional	taxon		x	
nameAccordingToID	optional	taxon		x	
namePublishedIn	optional	taxon		x	
namePublishedInYear	optional	taxon		x	
nomenclaturalCode	optional	taxon		x	
nomenclaturalStatus	optional	taxon		x	
originalNameUsage	recommended	taxon		x	
originalNameUsageID	recommended	taxon		x	
parentNameUsage	recommended	taxon		x	
parentNameUsageID	recommended	taxon		x	
phylum	recommended	taxon		x	
scientificNameAuthorship	recommended	taxon		x	
specificEpithet	recommended	taxon		x	
subgenus	recommended	taxon		x	
taxonConceptID	optional	taxon		x	
taxonID	optional	taxon		x	
taxonomicStatus	optional	taxon		x	
taxonRank	strongly recommended	taxon		x	
taxonRemarks	recommended	taxon		x	
verbatimTaxonRank	recommended	taxon		x	
vernacularName	recommended	taxon		x	
type or eventType	strongly recommended	event	x		
class	recommended	taxon		x	
family	recommended	taxon		x	

Term	OBIS Required	DarwinCore Class	Event	Occurrence eMoF	DNA
genus	strongly recommended	taxon		x	
kingdom	strongly recommended	taxon		x	
order	strongly recommended	taxon		x	

# **Part IV**

## **Ensuring Data Quality**

## 5 Data quality control

OBIS ignores records that do not meet a number of standards. For example, all species names need to be matched against an authoritative taxonomic register, such as the World Register of Marine Species. In addition, quality is checked against the OBIS required fields as well as against any impossible values. OBIS checks, rejects and reports the data quality back to the OBIS nodes, but never change records. The OBIS tier 2 nodes are responsible for the data quality and communicate errors back to the data providers. A number of QC tools are developed to help data providers and OBIS nodes:

- [QC tool for species names](#)
- [QC tool for geography and data format](#)

For specific concerns regarding quality control checks or issues, please submit a GitHub ticket to the [OBIS QC repository](#).

### 5.1 Why are records dropped?

Records can be dropped and therefore not published with your dataset for a number of reasons, including:

- The species is not marine
- The 'scientificName' or `scientificNameID` did not match with WoRMS
- Issues with coordinates:
  - No coordinates given
  - `decimalLatitude` or `decimalLongitude` out of range
- The coordinate is zero

For each dataset published, a quality report is generated where the number of dropped records and other quality issues will be flagged. Such reports can also be found when searching for data in OBIS. For example, if we searched for 'Crustacea' records, the following data quality report is given:

We can see that >110,222 Crustacean records have been dropped, mostly due to records missing coordinates or species being flagged as non-marine. Because species are determined as

## DATA QUALITY

### DROPPED RECORDS

Dropped records	114,372	
> Not marine	53,459	
> Marine unsure	31,502	
> No coordinates	57,780	
> Zero coordinates	11,328	

Figure 5.1: Number of Crustacean records dropped

being marine by WoRMS, we acknowledge that sometimes species are marked as `not_marine` erroneously. For specific advice on this topic, see the [common QC issues page](#).

To minimize the number of records dropped, be careful when formatting your data so that you are meeting the requirements.

## 5.2 How to conduct Quality Control

Once you have formatted your data for OBIS, or have received a formatted dataset, it is important to run quality control checks before publishing the dataset on the IPT. The following is a list of various tools you can use to help you perform quality checks on your data:

- R package [obistools](#)
- EMODnet Biocheck
  - [Web UI](#) built on obistools. This tool requires your dataset to be published on an IPT (e.g., a test IPT such as <https://ipt.gbif.org/> where your dataset will not be harvested by GBIF or OBIS). Note you are [required to have a login](#) to access an IPT
  - [R package](#)
- [Lifewatch data services](#)
- The US Integrated Ocean Observing System [Standardizing Marine Bio Data Guide](#)
- [WoRMS taxon match tool](#)



- [Other WoRMS web services, incl. taxon match](#)
- Excel Conditional Formatting tool
  - Excel > Home > Conditional Formating > Highlight cells Rules > Duplicate values...
- [GBIF data validator](#)
- [Python library for OBIS QC](#) developed by Canadian Integrated Ocean Observing System
- R package and function Hmisc:: describe
  - Can give important summary statistics and identify numbers that don't match

### 5.2.1 Conducting QC with obistools

To use `obistools` to conduct quality control, you can follow this general order:

1. Check that the taxa match with WoRMS
  - `obistools::match_taxa`
2. Check that all required fields are present in the occurrence table
  - `obistools::check_fields`
3. Check coordinates
  - Plot them on a map to identify any points that appear outside the scope of the dataset `obistools::plot_map`
  - Identify points with `obistools::identify_map`
  - Check that points are not on land `obistools::check_onland`
  - Ensure depth ranges are valid `obistools::check_depth`
4. Check for [statistical outliers](#) which may have had data entry errors
  - `obistools::check_outliers_species` and `obistools::check_outliers_dataset`
5. Check that the eventID and parentEventID are structured correctly `obistools::check_eventids`
  - Ensure all eventIDs in extensions have matching eventIDs in the core table `obistools::check_extension_eventids`
6. Check that eventDate is formatted properly `obistools::check_eventdate`

### 5.2.2 QC with R package Hmisc

The R package `Hmisc` has the function `describe` which can help you identify any discrepancies in your dataset.

It will summarize each of your variables for a given data field. This can help you quickly identify any missing data and ensure the number of unique IDs is correct. For example, in an Occurrence table with 1000 records, there should be 1000 unique occurrenceIDs.

```
library(Hmisc)
library(Hmisc)
data<-read.csv("example_data_occur.csv")
describe(data)
```

```
12 Variables      407 Observations
```

```
-----
CollectionCode
```

n	missing	distinct	value
407	0	1	BIOFUN1

```
Value      BIOFUN1
```

```
Frequency    407
```

```
Proportion    1
```

```
-----
eventID
```

n	missing	distinct
407	0	27

```
lowest : BIOFUN1_BF1A01 BIOFUN1_BF1A02 BIOFUN1_BF1A03 BIOFUN1_BF1A04 BIOFUN1_BF1A05
```

```
highest: BIOFUN1_BF1M3  BIOFUN1_BF1M4  BIOFUN1_BF1M6  BIOFUN1_BF1M8  BIOFUN1_BF1M9
```

```
-----
occurrenceID
```

n	missing	distinct
407	0	407

```
lowest : CSIC_BIOFUN1_1  CSIC_BIOFUN1_10  CSIC_BIOFUN1_100 CSIC_BIOFUN1_101 CSIC_BIOFUN1_102
```

```
highest: CSIC_BIOFUN1_95  CSIC_BIOFUN1_96  CSIC_BIOFUN1_97  CSIC_BIOFUN1_98  CSIC_BIOFUN1_99
```

This video shows how to use both `obistools` and `Hmisc` to conduct QC checks in R.

**Part V**

**Publishing Data**

## 6 Data publication and sharing

Once you have finished [formatting your data](#) and have conducted some basic [quality control procedures](#) on it, you are ready to publish your dataset to OBIS. Note that OBIS nodes can accept any data files from its data sources or data providers, and they publish these data on their respective OBIS node Integrated Publishing Toolkit (IPT). Data from [node IPTs](#) are then harvested by central OBIS to become [accessible via the global database](#). The Integrated Publishing Toolkit (IPT) is developed and maintained by the Global Biodiversity Information Facility (GBIF). While GBIF maintains an [IPT manual](#), we outline specific OBIS instructions here.

Some nodes require the **manager themselves to upload data**. This means after you have formatted your data for OBIS and have done some [basic quality control steps](#), you simply pass this file on to the appropriate node manager. They will communicate with you any issues during the publication process.

Sometimes a node manager will request the **data provider to upload their own dataset** to the IPT. If you the data provider are required to upload your data then there are a number of steps you will take to upload your data. A reminder that all data formatting and quality control steps should be completed before uploading your dataset. Once uploaded, the node manager will check your upload before publishing the dataset.

There are a few steps involved to publish a dataset, whether you are the provider or the node manager. You must:

1. Identify the correct [IPT](#) for your [OBIS node](#) (you may have to contact your node manager to confirm IPT)
2. Login to the IPT, or have the node manager create an account for you if you do not have one, so you can upload your dataset(s)
3. Map each of your fields to Darwin Core terms. This should be relatively straight forward if you have done this already during [data formatting](#)
4. [Fill all relevant metadata](#) to help users understand and cite your dataset
5. Publish and make your data public

Details for each of these steps are outlined on the subsequent pages. We will start with an overview of which Creative Commons license should be used, and then move on to using the IPT to publish data.

## 6.1 Licenses

OBIS nodes must make the necessary agreements with the original data providers so that data can be made available to OBIS under one of the following Creative Commons licenses (in order of preference):

- [CC0](#) - data may be used without restrictions
- [CC-BY](#) - data are available for any use if proper attribution and credit is given
- [CC-BY-NC](#) - data may be used for any non-commercial use as long as proper attribution/credit is given

You may need to consult with your organization if there are any copyright concerns. For more information on the different Creative commons license types see [About the licenses](#).

**Part VI**

**Access Data from OBIS**

## 7 Data access

OBIS has over 100 million records of marine data accessible for downloading. To download data from OBIS, there are several options:

- [OBIS homepage](#) or [advanced dataset search](#)
- [OBIS Mapper](#)
- Accessible through the [R package robis](#)
- [OBIS API](#)
- [Full data exports](#)
- [IPT](#)

**NOTE** When you download data from the Mapper or full export, the data you will receive is flattened into one table with occurrence plus event data. eMoF data tables are separate upon request. However when you download a dataset from the OBIS homepage or dataset page, all tables (Event, Occurrence, eMoF) are separate files.

### 7.1 OBIS Homepage and dataset pages

From the OBIS homepage, you can search for data in the search bar in the middle of the page. You can search by particular taxonomic groups, common names, dataset names, OBIS nodes, institute name, areas (e.g., Exclusive Economic Zone (EEZ)), or by the data provider's country.

When you search by dataset you will notice an additional option appears for [advanced search options](#). This will allow you to identify specific datasets, and apply filters for OBIS nodes and whether datasets include extensions.

Regardless if you found a dataset through the homepage or the advanced Dataset search, you will be able to navigate to individual dataset pages. For individual dataset pages (instead of aggregate pages for e.g., a Family) there are three buttons available:

- Report issue - allows you to report any issues with the dataset in question
- Source DwC-A - download the dataset as a Darwin Core-Archive file. This will provide all data tables as separate files within a zipped folder
- To mapper - this will open another browser with the data shown in the Mapper

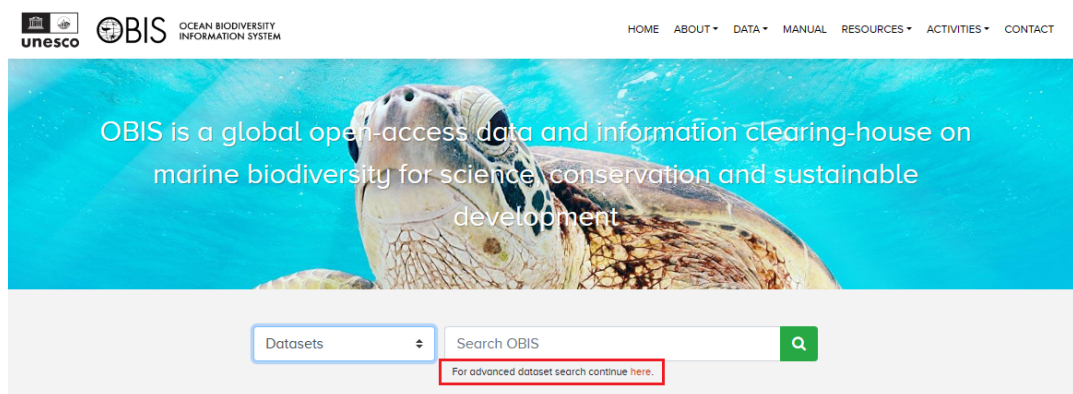


Figure 7.1: OBIS homepage search, showing where to find the advanced search link

### New Zealand research tagging database

URL	<a href="https://nzobisipt.niwa.co.nz/resource?r=mpl_tag">https://nzobisipt.niwa.co.nz/resource?r=mpl_tag</a>		
Repository URL	<a href="https://nzobisipt.niwa.co.nz/">https://nzobisipt.niwa.co.nz/</a>		
Node	SWP OBIS		
Published	2018-08-08 20:59		
Abstract	Tagging programmes have been used to provide information on fish and fisheries to central government policy makers in New Zealand for many years. A wide variety of species have been the subject of such studies, including finfish, shellfish and rock lobsters. In New Zealand, the Ministry for Primary Industries (formerly the Ministry of Fisheries) has funded these programmes to aid with fisheries research and stock assessment. Data from these programme are held in the "tag" database, from which the data in this dataset are sourced.		
Citation	Ministry for Primary Industries (2014). New Zealand research tagging database, Southwestern Pacific OBIS, National Institute of Water and Atmospheric Research (NIWA), Wellington, New Zealand, 411926 records, Online <a href="http://nzobisipt.niwa.co.nz/resource.do?r=mpl_tag">http://nzobisipt.niwa.co.nz/resource.do?r=mpl_tag</a> released on November 5, 2014.		
Rights	This work is licensed under a Creative Commons Attribution (CC-BY) 4.0 License		
Keywords	Occurrence, Observation		
Contacts	Creator	Kevin Mackay	
		NIWA	
	Contact	David Fisher	
		NIWA	
	Metadata Provider	Kevin Mackay	
		NIWA	
	Custodian Steward	David Fisher	

Darwin Core Archive as provided to OBIS by the OBIS node

[report issue](#)
[source DwC-A](#)
[to mapper](#)

Figure 7.2: Dataset download



If you searched for aggregate datasets (e.g., all Crustacea records, all records from OBIS-Canada, etc.), the **source** DwC-A button will not be available to you. To download these data subsets, you must click to **mapper** and then [download the data from the Mapper as a CSV](#).

## 7.2 Mapper

- <https://mapper.obis.org>

Watch this video demonstration of how to use the Mapper as well as the OBIS homepage search.

The mapper allows users to visualize and inspect subsets of OBIS data. A variety of filters are available (taxonomic, geographic, time, data quality) and multiple layers can be combined in a single view. Layers can be downloaded as CSV files.

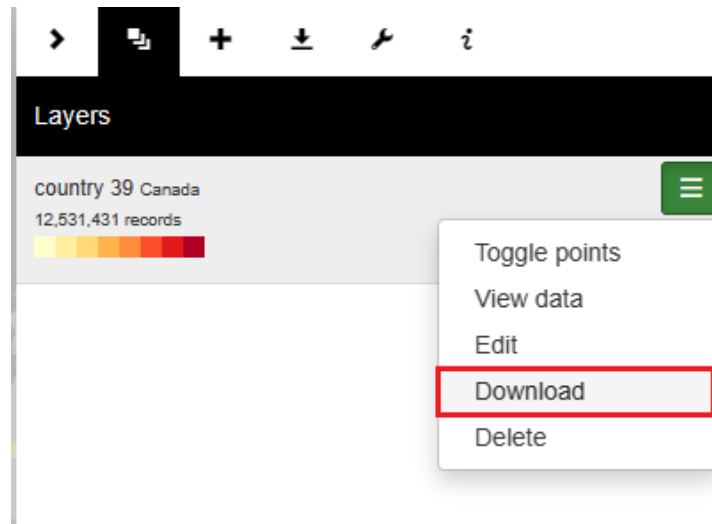


Figure 7.3: Screenshot demonstrating where how to download a particular layer

When you download data from the mapper, you will be given the option to include eMoF and/or DNA Derived Data extensions alongside the Event and Occurrence data. You must check the boxes of extensions you want to include in your download.

After downloading, you will notice that the Event and Occurrence data is flattened into one table, called “Occurrence.csv”. Upon inspecting this file in your viewer of choice, you will see it contains all 225 possible DwC fields, although not every field will contain data for each observation. Any extensions you checked will be downloaded as separate tables.

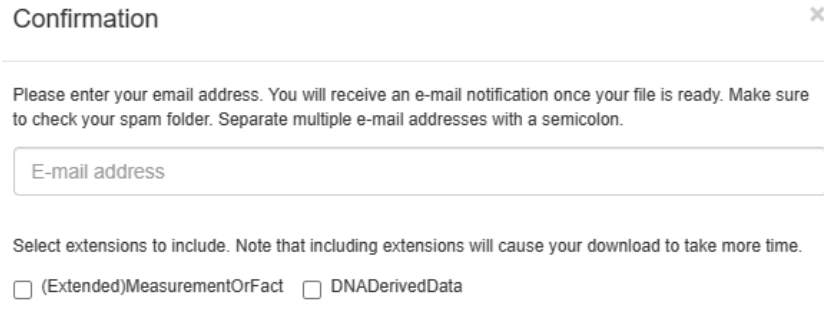


Figure 7.4: Screenshot showing the popup confirmation for which extensions you want to include in your download from the OBIS Mapper

## 7.3 R package

- <https://github.com/iobis/robis>

The robis R package has been developed to facilitate connecting to the OBIS API from R. The package can be installed [from CRAN](#) or [from GitHub](#) (latest development version). The package documentation includes a function reference as well as a [getting started vignette](#). As a quick example of what the package can do, you can obtain raw occurrence data by feeding a taxon name or AphiaID to the `occurrence` function.

If you'd like to then download this data, you can simply export R objects with the `write.csv` function. For example, if we wanted to obtain Mollusc data from OBIS:

```
library(robis)
moll<-occurrence("Mollusca")
write.csv(moll, "mollusca-obis.csv")
```

This file will be saved to your working directory (if you are not familiar with working directories, read [here](#)). After opening the file, you will notice that the fields in the download do not include every possible field, but instead only those where information has been recorded by data providers, plus the [fields added by OBIS's quality control pipeline](#).

To use robis for visualizing and mapping occurrences, see the [Visualization](#) section of the manual.

Watch the video below for a walkthrough of how to use the robis package to obtain OBIS data.

## 7.4 API

- <https://api.obis.org/>

Both the mapper and the R package are based on the [OBIS API](#), which can also be used to find and download data. When using the API directly, you can filter by the following options:

- Occurrence
- Taxon
- Checklist
- Node
- Dataset
- Institute
- Area
- Country
- Facet
- Statistics

When you have entered all the information you are interested in filtering by, scroll down and click the “Execute” button. This will produce a response detailing how many records match your criteria, as well as information for some of the headers from the data (e.g., basisOfRecord, Order, genus, etc.). A download button will be available for you to download the data as well.

When searching with the API, you may need to know certain identifiers, including:

- AphiaID - obtainable from the WoRMS page of a taxa of interest (e.g. the AphiaID for [Mollusca](#) would be 51)
- Dataset UUID - can be obtained from the URL on individual dataset pages
  - E.g., [this dataset's](#) UUID would be 5061d21c-6161-4ea2-a8d4-38f8285dfc47
- Area ID
- Institute ID - this should be the Ocean Expert ID (e.g., the ID for [NOAA Fisheries Service, Southeast Regional Office St. Petersburg](#) is 7532)
- OBIS node UUID

A short video demonstrating use of the API is shown below.

## 7.5 Full exports

- <https://obis.org/data/access/>

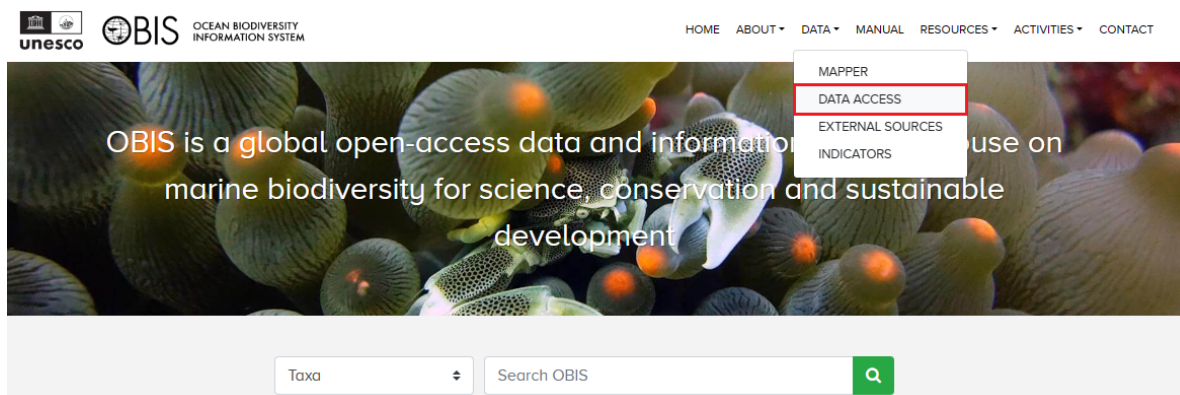


Figure 7.5: OBIS homepage showing where to navigate to access full database exports

To obtain a full export of OBIS data, navigate to the OBIS homepage, click on Data from the top navigation bar, then select [Data Access](#) from the dropdown menu.

Here you will be able to download all occurrence records as a CSV or Parquet file. Note the disclaimer that such exports will not include measurement data, dropped records, or absence records. As with downloads from the Mapper, the exported file is a single Occurrence table. This table includes all provided Event and Occurrence data, as well as 68 fields added by the OBIS Quality Control Pipeline, including taxonomic information obtained from WoRMS.

## Data access

OBIS harvests occurrence records from thousands of datasets and makes them available as a single integrated dataset. [Read more about data access in the OBIS manual.](#)

### Full exports

We provide periodic exports of the entire set of quality controlled presence records as CSV. This is the easiest way to download data for large scale analyses. **Absence records, records of insufficient quality, or measurements records are not included. Use the API or R package to access these data or contact [p.provoost@unesco.org](mailto:p.provoost@unesco.org).**

Full OBIS export 2023-02-08	csv	<a href="#">download</a>
Full OBIS export 2023-02-08	parquet	<a href="#">download</a>

Each export consists of a single occurrence table which includes all Darwin Core fields provided by the data providers as either Event or Occurrence records, with the exception of the fields below which have been transformed or added by the OBIS quality control pipeline.

Figure 7.6: OBIS Data Access page

## 7.6 Finding your own data in OBIS

To find your own dataset in OBIS, you can use the same tools as finding any dataset in OBIS. You have the following options:

- From the [OBIS homepage](#) or the [Mapper](#), you can search by dataset name, species of interest, the OBIS node that you uploaded to, or by institute
  - Note: When using the Mapper you can combine multiple search criteria to help narrow down your search
    - \* E.g., if we wanted to find [this dataset](#) in the Mapper, we could search for OBIS USA under Nodes, National Oceanic and Atmospheric Administration, Washington under Institutes, and/or Radioczoa under Scientific Name. Then when we view the data and scroll down to datasets, the only one listed is the one we were interested in
- If you have used the (extended)measurementOrFact extension and have measurement-Type data, you can [search by the name of your measurementType](#), click on the hyperlink for records. This will populate a list of datasets that you can scroll through.

## 7.7 How to contact data provider

To contact the data provider, navigate to the page for the individual dataset in question (e.g., <https://obis.org/dataset/80479e14-2730-436d-aaaa-b63bdc7dd06f>). Under the “Contacts” section, there will be a list of individuals you can contact. Clicking any name will direct you to your system’s default email program. For example:

Contacts	Creator	<b>Todd O'Brien</b> National Oceanic and Atmospheric Administration
	Contact	<b>Todd O'Brien</b> National Oceanic and Atmospheric Administration
	Metadata Provider	<b>Abby Benson</b> U.S. Geological Survey
	Publisher	<b>Abby Benson</b> U.S. Geological Survey

Figure 7.7: *Example of contact section on a dataset homepage access via the OBIS search*

If you are the node manager and need to contact the data provider about a particular dataset, contact information should be provided in the metadata and you can contact them from information provided.

## 7.8 Interpreting downloaded files from OBIS

In general, the field names you will see when you download data from OBIS are the same as those seen during the data formatting and publishing process. When you download data from

the [Mapper](#) you will see all 225 possible Darwin Core fields.

Downloading data from an IPT or full export will include only the fields provided by the data provider, formatted as one Occurrence file (or separate files for individual datasets). Some fields are added through the OBIS quality control pipeline, including taxonomic information from WoRMS and the fields **flags**, **bathymetry**, and **dropped**. As mentioned in the [Quality Control section](#), the fields **flags** and **dropped** will list quality control issues or if the record was dropped, respectively. Details and definitions for all fields added by the OBIS QC pipeline can be found [here](#).

For a full list of the other Darwin Core terms and their definitions included in downloads, please reference the [Darwin Core reference guide](#).

## **Part VII**

# **Data Visualization and Analysis**

## 8 Data Visualization

### 8.1 Example notebooks using data from OBIS

Here are a few R notebooks showcasing the robis package:

- [Data exploration of wind farm monitoring datasets in OBIS](#)
- [Diversity of fish and vulnerable species in Marine World Heritage Sites based on OBIS data](#)
- [Data exploration - Stratified random surveys \(StRS\) of reef fish in the U.S. Pacific Islands](#)
- [DNADerivedData extension data access](#)
- [Canary Current LME](#)

Here are others that may be of interest:

- [Diversity indicators using OBIS data](#)
- [OBIS species richness for OSPAR](#)
- [Quality control of ISA data](#)
- [Accessing gridded data](#)

### 8.2 obisindicators: calculating & visualizing spatial biodiversity using data from OBIS

[obisindicators](#) is an R library developed during the [2022 IOOS Code Sprint](#). The purpose was to create an ES50 diversity index within hexagonal grids following the [diversity indicators notebook](#) by Pieter Provoost linked above. The package includes several examples, limited to 1M occurrences, that demonstrate uses of the package.



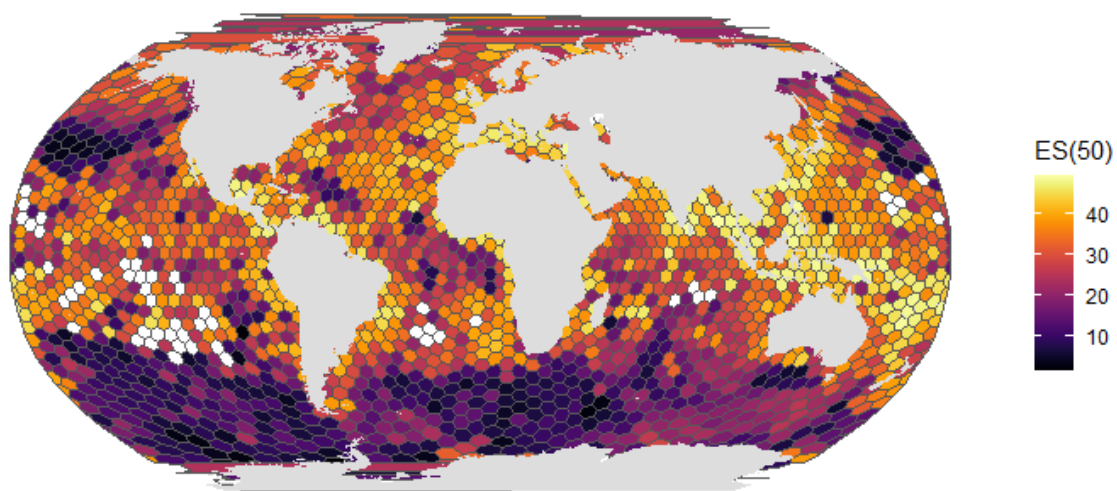


Figure 8.1: screenshot

**Part VIII**

**Additional Resources**

## 9 Other Resources

In this section we highlight useful resources created by collaborators and other community members.

### 9.1 MBON Pole to Pole Tutorial

- <https://www.youtube.com/watch?v=teJhfsSWonE>

This tutorial was created by the [MBON Pole to Pole project](#) to help guide people through the process of transforming datasets to Darwin Core using [tools](#) MBON Pole to Pole has developed.

### 9.2 IOOS Darwin Core Guide

- [https://ioos.github.io/bio\\_data\\_guide/](https://ioos.github.io/bio_data_guide/)

This book contains a collection of examples and resources related to mobilizing marine biological data to the [Darwin Core standard](#) for sharing through [OBIS](#). This book has been developed by the [Standardizing Marine Biological Data Working Group \(SMBD\)](#). The working group is an open community of practitioners, experts, and scientists looking to learn and educate the community on standardizing and sharing marine biological data.

### 9.3 EMODnet Biology

- <https://classroom.oceanteacher.org/course/view.php?id=430>

Contributing Datasets to EMODnet Biology is a course hosted on [Ocean Teacher Global Academy \(OTGA\)](#), developed by members of the [European Marine Observation and Data Network](#). The course prepares users to format, publish, and perform quality control checks on datasets according to Darwin Core standards. While targeted at EMODnet Biology users, this course has significant overlap in how to prepare datasets for OBIS and is useful for those unfamiliar with OBIS standards. Note, an account with OTGA is required to access the course.

## 9.4 Template Generators

There is an [Excel template generator](#) developed by Luke Marsden & Olaf Schneider as part of the Nansen Legacy project. It allows the creation of Event or Occurrence core templates, with an optional eMoF extension. Note this template generator is aimed at GBIF users, so make sure to account for and include required OBIS terms.

There is also an [Excel to Darwin Core macro tool](#) developed by GBIF Norway that you can download for use in Microsoft Excel. This macro can help you set up Event, Occurrence, and eMoF tables by selecting all relevant DwC fields from a list, or by importing data from another spreadsheet. It allows for auto-generation of identifiers (e.g. eventID, occurrenceID) if macros are enabled, and can also auto-populate the eMoF when measurement fields in the Occurrence table are populated.