

OBIS: Incorporation of eDNA data

Mathew Biddle, Stephen Formel

Tuesday, June 4th 2024

3rd National Workshop on Marine eDNA

John Hopkins Applied Physics Laboratory

Laurel, Maryland



Resources!!

Background	Training	Guidance	Community Support
<u>OBIS</u>	<u>Data Formatting for OBIS: Video Playlist</u>	<u>OBIS Manual</u>	<u>Standardizing Marine Biological Data Working Group (SMBD)</u>
<u>OBIS-USA</u>	<u>OceanTeacher Global Academy (OTGA): Contributing and publishing datasets to OBIS</u>	<u>Guide: Publishing DNA-derived data through biodiversity data platforms</u>	Slack: <u>SMBD</u> Slack: <u>OBIS Slack</u> Slack: <u>GBIF-NA</u>
<u>GBIF</u>	<u>GBIF: Training and Learning Resources</u>	<u>IOOS Bio Data Guide</u>	<u>GBIF Discourse</u>
<u>GBIF-US</u>	<u>OBIS Genetic Data Webinar (video)</u>	<u>GBIF IPT Manual</u>	<u>obis-usa@usgs.gov</u> <u>gbif-us@usgs.gov</u>
<u>GBIF-North America</u>	<u>GBIF Tech Support Hour: DNA data publishing (video)</u>	<u>NOAA Omics Data Management Guide</u> <u>EMO BON Data Management Plan</u>	<u>helpdesk@obis.org</u> <u>helpdesk@gbif.org</u>
<u>GBIF and DNA</u>	<u>TOPs: Open Science 101</u>	The <u>DwC Quick Reference</u>	

OBIS? GBIF? What's the diff?

GBIF.us About ▾ News ▾ Data Literature

June 3, 2024

OBIS and GBIF endorse joint strategy and action plan for marine biodiversity data



GBIF = all biodiversity data

OBIS = ocean biodiversity data

There are simple tools available to publish to both simultaneously.

Both built on same (meta)data model

<https://www.gbif.us/post/2024/gbif-obis-joint-action-plan/>

- International initiative that aims to provide open access to biodiversity occurrence data.
- Most well-known for their publishing platform, and their role as data aggregator
- Represent an investment from the international political community and a vibrant scientific community of nodes, publishers, and users of standards and practices.



Occurrence Records: 126,826,047

Datasets: 4,856

Taxa: 149,319



USGS manages the US node: [OBIS-USA](https://www.usgs.gov/obis-usa)

- Represent US interests in the international initiatives
- Advise the US scientific community on biodiversity informatics, data, and standards
- Mobilize and synthesize data to enhance its FAIRness and relevance on national and international scales.
- Archive at NOAA NCEI and meets PARR Requirements

The screenshot shows the OBIS USA website interface. At the top, there are logos for UNESCO and OBIS (Ocean Biodiversity Information System) and a navigation menu with links for HOME, ABOUT, DATA, MANUAL, RESOURCES, ACTIVITIES, and CONTACT. The main heading is "OBIS USA".

Feeds: A list of RSS feeds from various institutions, including ichthyology, vertnet, gbif, bison.usgs.gov, nhm.ku.edu, idigbio, geome-db, calacademy, and flmnh.ufl.edu.

Description: Ocean Biodiversity Information System USA (OBIS-USA) brings together marine biological observation data – recorded observations of identifiable marine species at a known time and place, collected primarily from U.S. Waters or with U.S. funding.

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

Contacts: Stephen Formel (sformel@usgs.gov), Abby Benson (on detail to the US National Nature Assessment until Sep 2025), Kyla Richards (krichards@usgs.gov), and Mathew Biddle (mathew.biddle@noaa.gov).

STATISTICS:

Occurrence records	12,848,644
> Species level	9,156,371
Absence records	12,579,612
Species	43,559
Taxa	58,332
Datasets	192
Time range	1700 - 2024

DISTRIBUTION: A world map showing the geographic distribution of marine biological observations, with a high density of red dots in the North Atlantic and Pacific Oceans.

OBIS and Darwin Core

- OBIS uses Darwin Core
- Darwin Core is a vocabulary standard for transmitting information about biodiversity
- Darwin Core follows a star schema
- Darwin Core has extensions

The screenshot shows an Excel spreadsheet with columns for Darwin Core terms. The data includes occurrence IDs, event dates, localities, and depth measurements. A large 'CSV' watermark is overlaid on the right side of the spreadsheet.

The Darwin Core Archive (DwC-A)

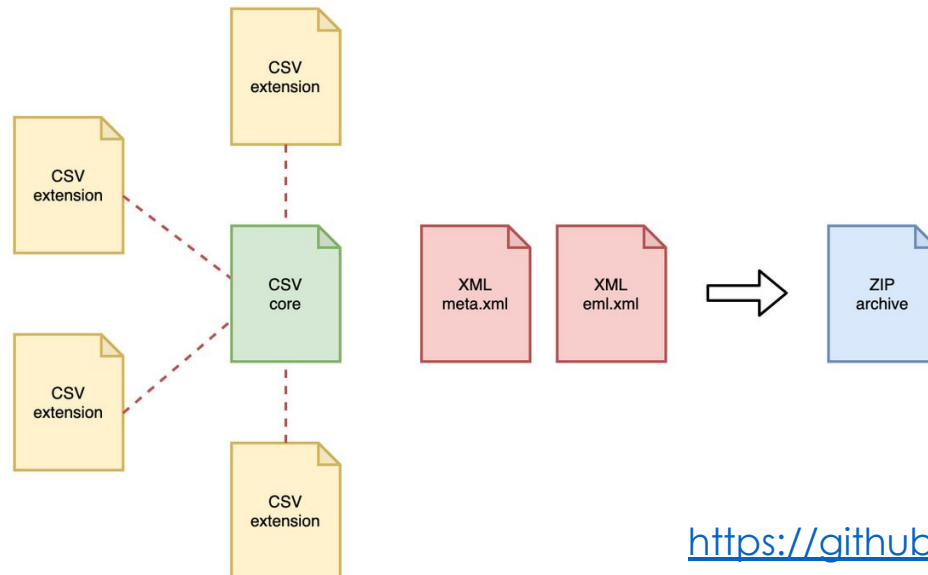


Table 1: Example Occurrence Data From Silliman et al., 2023

Term	Example
occurrenceID	GOMECC4_PANAMACITY_Sta21_DCM_A_occ7c8b2f5e16137114160dfd4001f67550
eventDate	2021-09-20T18:04-04:00
locality	USA: Gulf of Mexico
locationID	PANAMACITY_Sta21
decimalLatitude	29.206
decimalLongitude	-85.647
geodeticDatum	WGS84
minimumDepthInMeters	39
maximumDepthInMeters	39

<https://github.com/tdwg/dwc>

Table 2: Example Occurrence Data From [Silliman et al., 2023](#)

Term	Example
occurrenceID	GOMECC4_PANAMACITY_Sta21_DCM_A_occ7c8b2f5e16137114160dfd4001f67550
basisOfRecord	MaterialSample
organismQuantity	35
organismQuantityType	DNA sequence reads
sampleSizeValue	12436
sampleSizeUnit	DNA sequence reads
associatedSequences	https://www.ncbi.nlm.nih.gov/sra/SRR26161072 https://www.ncbi.nlm.nih.gov/biosample/SAMN37516159 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA887898
identificationRemarks	Tourmaline; qiime2-2021.2; naive-bayes classifier, confidence (at lowest specified taxon): 0.960018364, against reference database: PR2 v5.0.1; V9 1391f-1510r region; 10.5281/zenodo.8392706. The PR2 database used for taxonomic assignment is primarily curated for protists, and may not accurately resolve metazoa, land plants or macrosporic fungi to lower taxonomic levels.
verbatimIdentification	Karenia brevis
scientificName	Karenia brevis
scientificNameID	urn:lsid:marinespecies.org:taxname:2330

Table 3: Example Occurrence Data (ambiguous taxonomy) From [Silliman et al., 2023](#)

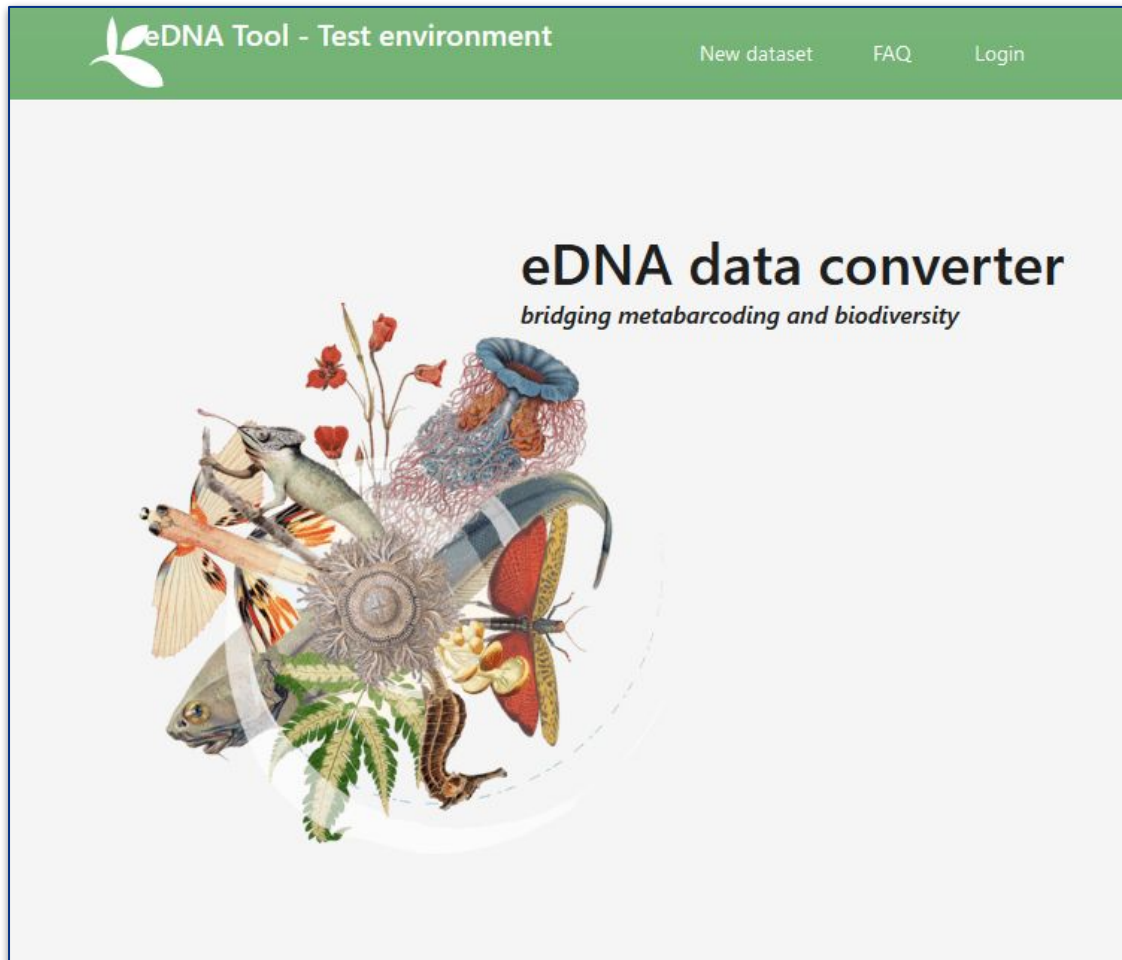
Term	Example
occurrenceID	GOMECC4_YUCATAN_Sta100_Surface_B_occ1f111363da96fee3d180ddb12741d4ce
basisOfRecord	MaterialSample
organismQuantity	18
organismQuantityType	DNA sequence reads
sampleSizeValue	12151
sampleSizeUnit	DNA sequence reads
associatedSequences	https://www.ncbi.nlm.nih.gov/sra/SRR26160967 https://www.ncbi.nlm.nih.gov/biosample/SAMN37516435 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA887898
identificationRemarks	Tourmaline; qiime2-2021.2; naive-bayes classifier, confidence (at lowest specified taxon): 0.960018364, against reference database: PR2 v5.0.1; V9 1391f-1510r region; 10.5281/zenodo.8392706. The PR2 database used for taxonomic assignment is primarily curated for protists, and may not accurately resolve metazoa, land plants or macrosporic fungi to lower taxonomic levels.
verbatimIdentification	Unassigned
scientificName	Biota incertae sedis
scientificNameID	urn:lsid:marinespecies.org:taxname:12

Examples

id	samp_nam	env_broad	env_local	env_medium	source_m	samp_coll	samp_mat	size_frac	samp_vol	nucl_acid	nucl_acid	lib_layout	target_gene	target_sul	pcr_cond	seq_meth	otu_class	otu_seq_c	otu_db	pcr_prime	pcr_prime	pcr_prime	pcr_prime	pcr_prime	DNA_seq_c
GOMECC4	GOMECC4	marine bio	marine m	sea water	GOMECC4	Niskin bot	Pumped t	0.22 µm	1920 ml	https://gi	10.1111/1	paired	16S rRNA	V4-V5	initial der	Illumina N	Tourmalin	Tourmalin	Silva SSU	IGTGYCAG	(CCGYCAA	515F-Y	926R	10.1111/1	TACGAGG

Table 4: Example Occurrence Data From [Silliman et al., 2023](#)

Term	Example
occurrenceID	GOMECC4_PANAMACITY_Sta21_DCM_A_occ7c8b2f5e16137114160dfd4001f67550
DNA_sequence	GCTCCTACCGATTGAGTGATCCGGTGAATAATTCGGACTGCCGCAGTGTTTCAGATCCTGAACGTTGCAGTGGAAAGTT
concentration	1.177
target_gene	18S rRNA
target_subfragment	V9
pcr_primer_forward	GTACACACCGCCCGTC
pcr_primer_reverse	TGATCCTTCTGCAGGTTACCTAC
pcr_primer_name_forward	1391f
pcr_primer_name_reverse	EukBr
pcr_primer_reference	10.1371/journal.pone.0006372
seq_meth	Illumina MiSeq 2x250
otu_class_appr	Tourmaline; qiime2-2021.2; dada2; ASV
otu_db	PR2 v5.0.1; V9 1391f-1510r region; 10.5281/zenodo.8392706
otu_seq_comp_appr	Tourmaline; qiime2-2021.2; naive-bayes classifier



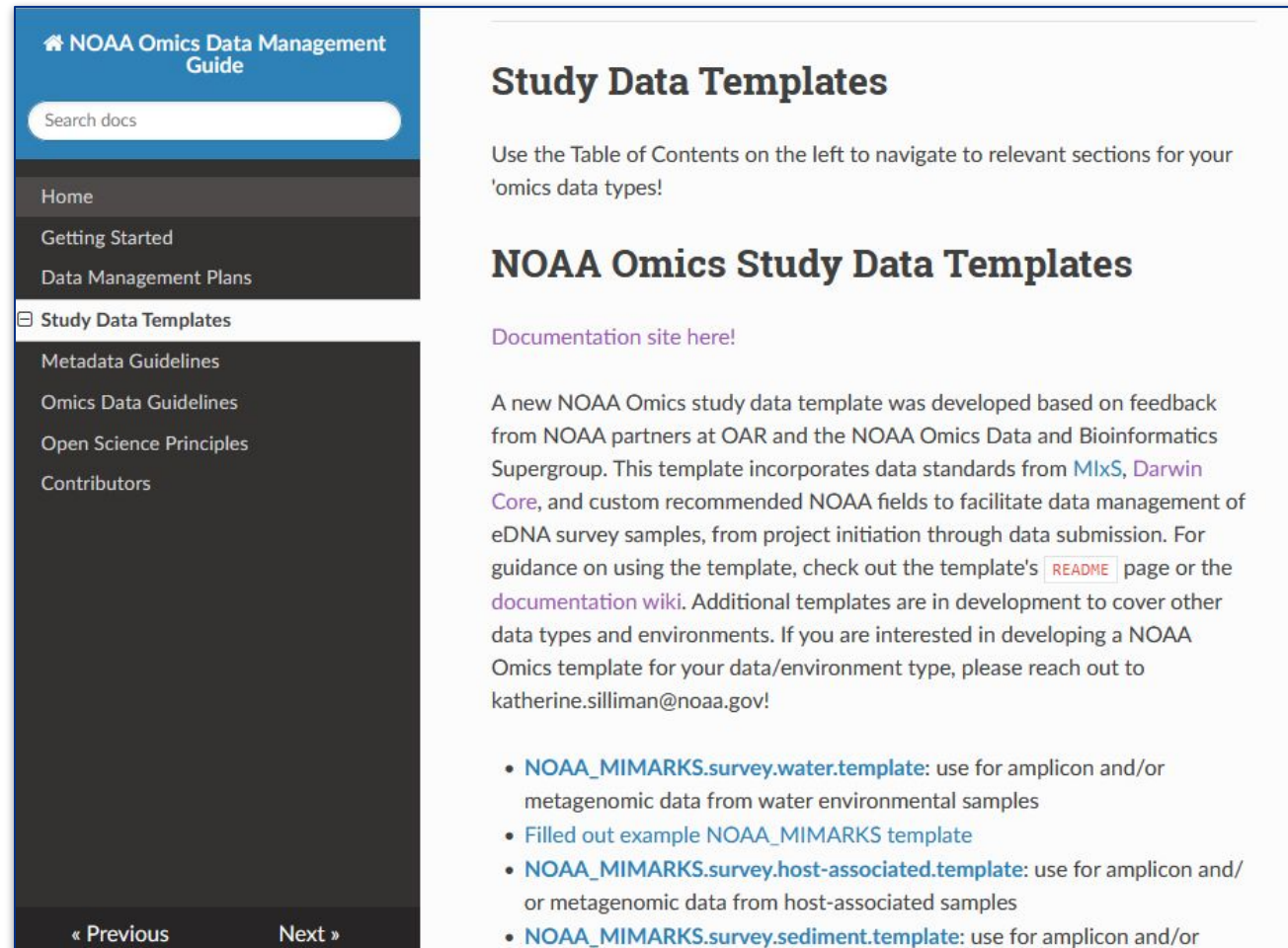
eDNA Tool - Test environment

New dataset FAQ Login

eDNA data converter

bridging metabarcoding and biodiversity

<https://edna-tool.gbif-uat.org/>



NOAA Omics Data Management Guide

Search docs

Home
Getting Started
Data Management Plans
Study Data Templates
Metadata Guidelines
Omics Data Guidelines
Open Science Principles
Contributors

Study Data Templates

Use the Table of Contents on the left to navigate to relevant sections for your 'omics data types!

NOAA Omics Study Data Templates

[Documentation site here!](#)

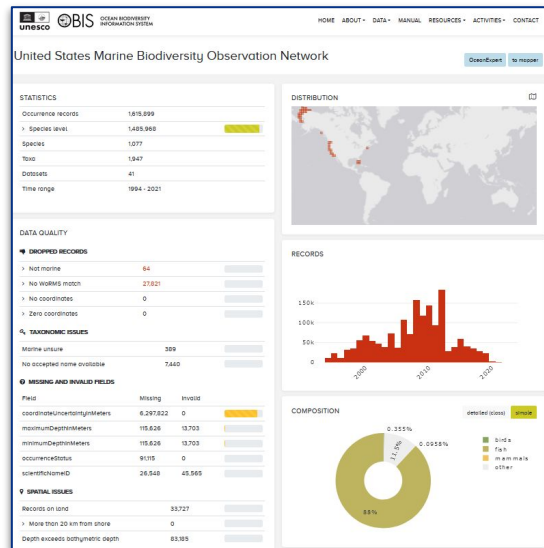
A new NOAA Omics study data template was developed based on feedback from NOAA partners at OAR and the NOAA Omics Data and Bioinformatics Supergroup. This template incorporates data standards from [MlxS](#), [Darwin Core](#), and custom recommended NOAA fields to facilitate data management of eDNA survey samples, from project initiation through data submission. For guidance on using the template, check out the template's [README](#) page or the [documentation wiki](#). Additional templates are in development to cover other data types and environments. If you are interested in developing a NOAA Omics template for your data/environment type, please reach out to katherine.silliman@noaa.gov!

- [NOAA_MIMARKS.survey.water.template](#): use for amplicon and/or metagenomic data from water environmental samples
- [Filled out example NOAA_MIMARKS template](#)
- [NOAA_MIMARKS.survey.host-associated.template](#): use for amplicon and/or metagenomic data from host-associated samples
- [NOAA_MIMARKS.survey.sediment.template](#): use for amplicon and/or

« Previous Next »

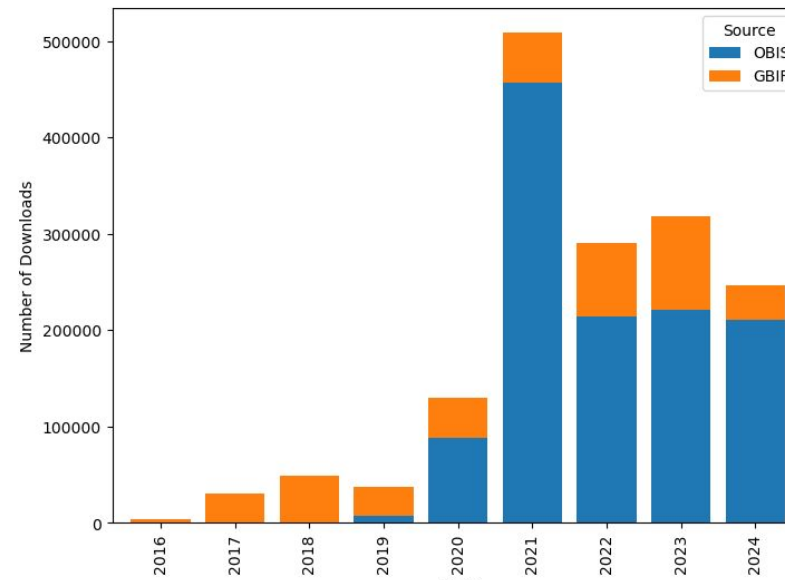
<https://noaa-omics-dmg.readthedocs.io/en/latest/study-data-templates.html>

Data use beyond original intent: U.S. MBON example



Observing methods include:

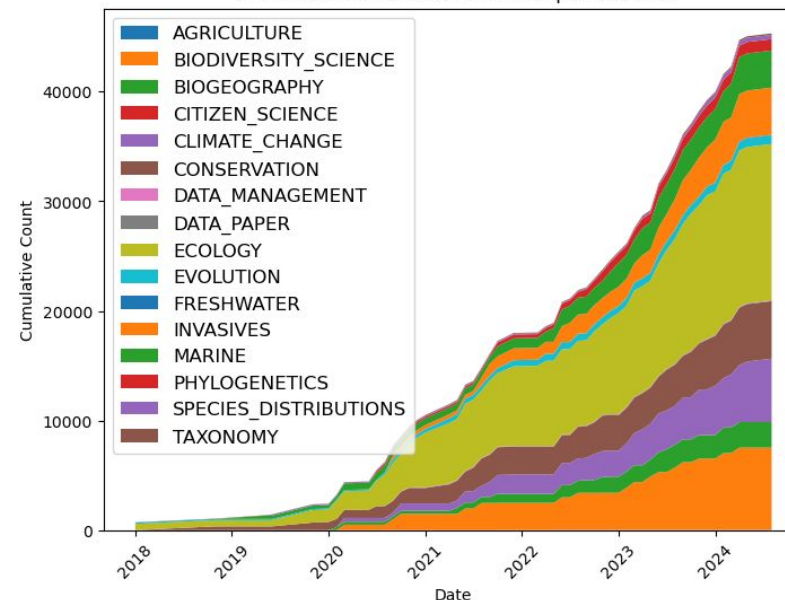
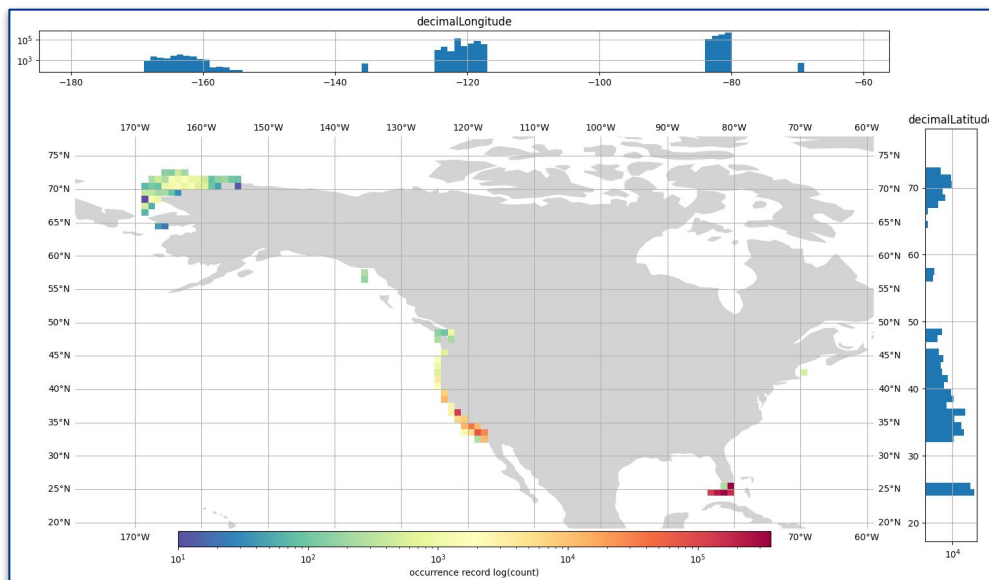
- diver survey
- fisheries indep. mon.
- **metabarcoding**
- photo plot
- transect survey
- trawl survey
- bongo net
- quadrat survey



Downloads of US MBON data

- OBIS: 1,198,306
- GBIF: 414,685

MBON Dataset Contributions to Topic Citations



Citations from GBIF

- 41 dataset DOI's
- 2,428 total citations

https://ioos.github.io/ioos_code_lab/content/code_gallery/data_access_notebooks/2022-11-23_pyobis_example.htm

Thank you!

Mathew Biddle

Mathew.Biddle@noaa.gov

ORCID: [0000-0003-4897-1669](https://orcid.org/0000-0003-4897-1669)

Stephen Formel

sformel@usgs.gov

ORCID: [0000-0001-7418-1244](https://orcid.org/0000-0001-7418-1244)

