

MBON Data Management and Cyberinfrastructure Working Group (MBON DMAC WG)

Mathew Biddle and Abby Benson
Wednesday , November 2, 2022

mathew.biddle@noaa.gov

albenson@usgs.gov



Purpose

WG purpose:

To implement interoperability, data archival, and publication standards for marine and coastal biodiversity observations.

Convergence on best practices for:

- Data formatting (from collection to application)
- Data interoperability
- Data archive
- Interconnectedness of databases

WG Structure/Expectations

- Monthly meetings (1:00 – 1:50 pm ET on the second Tuesday)
 - First 10 mins to review dataset progress
 - Announcements - upcoming meetings/related activities
 - Selected topic (eg. ATN, EDI,
- Participation and engagement from each project
 - Where are the trouble spots?
 - Where can this group help reduce those barriers?
- Coordination with other MBON working groups (mini teams)
 - eDNA, animal tracking, sound

What have we done so far?

MBON dataset registration

Please register each individual dataset that has started collecting observations.

Scope of registration:
This registration form collects information about datasets that have started collecting data.
We will use the information collected here to evaluate where each dataset is along the data management and prioritization archival process.

Darwin Core Guide

Preface

1 Introduction

2 Applications

2.1 Aligning Data to Darwin Core ...

2.2 Salmon Ocean Ecology Data

2.3 Hakai Seagrass

2.4 Trawl Data

2.5 dataset-edna

Appendix

A Frequently Asked Questions

B Tools

B.1 R

B.2 Python

B.3 Google Sheets

Updated: 2022-06-17

Darwin Core Guide

By: Standardizing Marine Biological Data Working Group



Preface

This book contains a collection of example resources related to mobilizing marine biological data using the Darwin Core standard.

12 new/updated layers in
MBON Portal ([link](#))

Home Code of Conduct Setup Episodes Extras License Improve this page 

Search...

Marine Biological Data Mobilization Workshop 2022

OBIS-USA, IOOS, Hakai, CIOOS, MBON, OTN

Online

March 14-15, 2022

13:00 - 17:00 EDT

Instructors: Abby Benson, Mathew Biddle, Sarah Bingo, Enrique Montes, Tylar Murray, Jonathan Pye, Tim van der Stap

Helpers: Ben Best, Brett Johnson, Dan Otis, Carolina Peralta

IOOS | Integrated Ocean Observing System

MBON Data and File Formatting Home

Menu

- Introduction
- MBON Data Flow
- Data and File Formatting
- Metadata and Documentation
- MBON use case

MBON Data and File Formatting

Summary: This documentation describes the Marine Biodiversity Observation Network (MBON) data and file formatting recommendations.

[Edit me](#)

Introduction

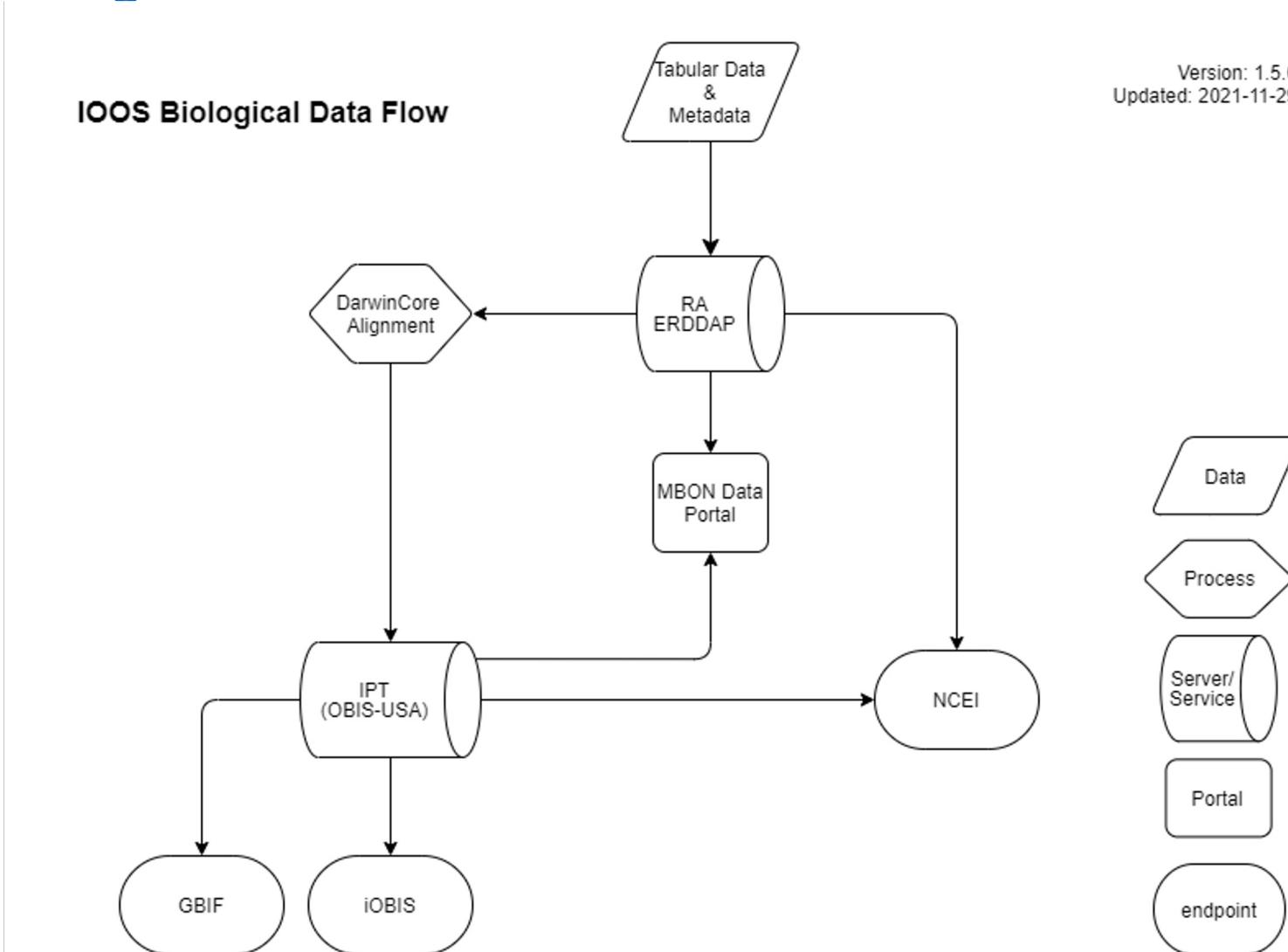
Marine Biodiversity Observation Network (MBON) observation data is focused on organisms from microbes to whales, including measures of biodiversity (e.g. presence, abundance), productivity, genomics, phenology, and other relevant ecological process measurements or indices. Also featured are habitat characterization and habitat diversity measures, including satellite data and added-value data derived from satellite observations, and neural network model results, such as biogeographical seascape classifications.

The data have been generated within the MBON regions of the Arctic, Central California, Southern California, the Gulf of Maine, the Pacific Northwest, and South Florida. Data have been collected by associated scientists or provided by multiple other independent programs, such as the IOOS Regional Associations, Long-Term Ecological Research (LTER) programs, universities, and other fisheries

 EYES ON THE OCEAN™

Data flows for project data

<https://ioos.github.io/mbon-docs/>



CREMP Dry Tortugas (South Florida MBON)

ERDDAP > search

ioos / bio_data_guide

Code Issues 18 Pull requests

master bio_data_guide / Standardized write_cremp_coral_csv_occurrence

mstoessel Add files via upload ...

1 contributor

1197 lines (1140 sloc) | 69.9 KB

```
1 %=====
2 % NAME
3 % write_fl_cremp_coral_occurrence_csv.m
4 % PURPOSE:
5 % The purpose of this program read and
6 % convert data to proper units and format
7 % HISTORY:
8 % Marion Stoessel, April 2017
9 % Department of Oceanography
10 % Texas A&M University
11 % Copyright TAMU
12 % Modified:
13 % July 2019, January 2021
14 %=====
```

bio_data_guide / Standardized write_cremp_coral_csv_occurrence

URL <https://www1.usgs.gov/obis-usa/ipt/resource?r=crempdrytortugas2012>

IOOS | Integrated Ocean Observing System

Marine Biodiversity Observation Network

All 35,039 Sensor Stations 34,688 Data Layers 842 Project data 33 Moving platforms 2 Variable Types 136 Affiliates 238

Florida Coral Reef Evaluation & Monitoring Project (CREMP)

The Coral Reef Evaluation and Monitoring Project (CREMP) was established in 1996 with the primary goal of examining temporal changes in coral benthic communities in the FKNMS and to provide information to management organizations on the status of coral reef resources. Reports provided by CREMP have assisted in evaluating the efficacy of the SPAs and have aided in the permit review process for coastal construction projects and collection of corals in the FKNMS. Since CREMP's inception in 1996, the CREMP methods and data collection efforts have expanded to apply to the Dry Tortugas National Park in the DRTO CREMP project that began in 1999 and to the coral reefs along Florida's southeast coast in the SECREMP project starting in 2003. Since monitoring was initiated, CREMP has documented changes in benthic community composition due to both large-scale and localized disturbances in the environment. While it's difficult to distinguish what proportion of the changes observed by CREMP is directly due to present day stressors as opposed to the anthropogenic disturbances propagated over the last century, CREMP has predominately observed a decline in the cover and diversity of corals in the FKNMS. These data describe the long-term trends observed by CREMP, DRTO CREMP, and SECREMP projects and provide insight as to how benthic community structure along the Florida reef tract has changed since 1996 and may be shaped in the future.

Time span Dec 31, 1995 10:00 (EST) - Jan 2, 2016 10:00 (EST)

Extent: 24.4517°,-83.0022° × 27.1317°,-80.0158°

Layers 1

CREMP_Dry_Tortugas_Yearly_Revisited_DATA_v3_2020

raw data + Matlab code + OBIS dataset + MBON data layer

of selected reefs in the Florida Keys National Marine collected provides information on the temporal changes continue to be underwater videography and timed from 1996 through 2008 in the Florida Keys and 1999 the percent cover of major benthic taxa (stony corals, sponges. Additionally, it examines the long-term trends of (e.g., and Porites astreoides) and the clionid sponge, extremely important in documenting the temporal

to the Public Domain (CC0 1.0)

Authorship_information, accepted_name_usage, citation_number, area, array, array-data, graphic, biology, class, commission, common,

early_revisited_DATA_v3_2013

early_revisited_DATA_v3_2014

early_revisited_DATA_v3_2015

it_CREMP_yearly_revisited_DATA_v3_2016

it_CREMP_yearly_revisited_DATA_v3_2017

it_CREMP_yearly_revisited_DATA_v3_2018

it_CREMP_yearly_revisited_DATA_v3_2019

it_CREMP_yearly_revisited_DATA_v3_2020

Marion Stoessel
(Texas A&M/GCOOS/SF
MBON)

[raw data](#)
+
[Matlab code](#)
+
[OBIS dataset](#)
+
[MBON data layer](#)

Wilkinson Basin Time Series Station (WBTS): Mesozooplankton 2004-2017

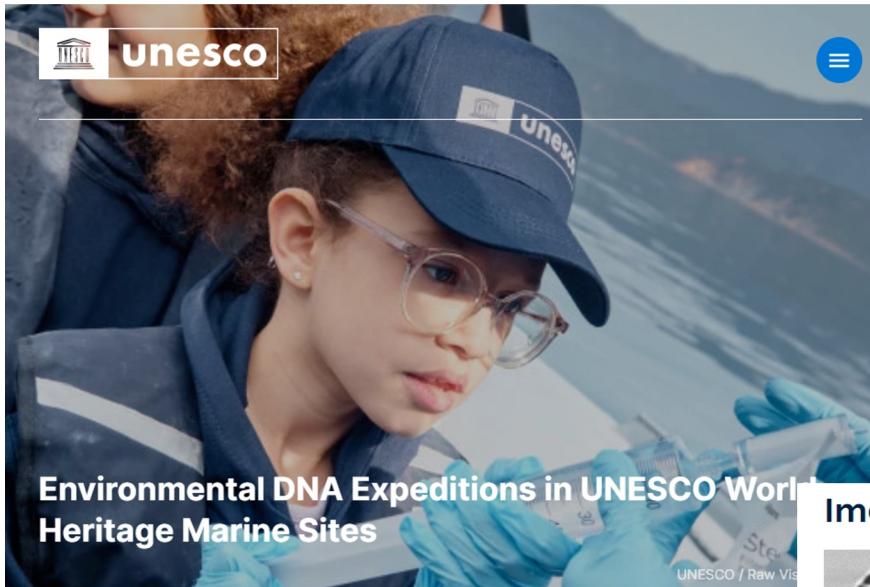
The collage consists of four overlapping screenshots:

- Top Left:** A GitHub repository page for `ioos/bio_data_guide`. It shows a pull request by Dylan-Pugh titled "Aligning Data or Fact using MBON-supp". The commit message says "Row enumeration now only processes valid counts". The repository has 1 contributor and 829 lines of code.
- Top Center:** The **MBON Data Portal** homepage. It features a large image of anemones and includes links to "EXPLORE MAP" and "SEARCH 270+ DATASETS".
- Bottom Left:** A Jupyter Notebook titled "Aligning Data or Fact using MBON-supp". It contains sections like "Ingest display", "How to use the portal", and "The steps". It also includes a screenshot of the MBON Data Portal interface.
- Bottom Right:** A detailed dataset page from the OBIS-USA Integrated Publishing Toolkit (IPT). It shows a table of data with columns for species names, counts, and coordinates, and a note about Darwin Core Archive file specification.

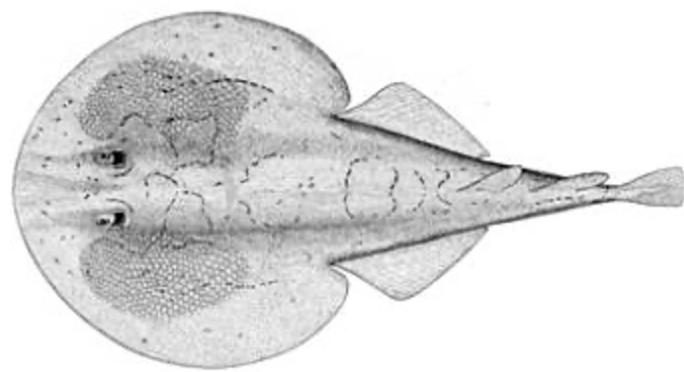
Dylan Pugh
(GMRI/NERACOOS/Gulf of Maine
MBON)

raw data
+
[Jupyter Notebook](#)
+
[OBIS dataset](#), [GBIF dataset](#), [NCEI dataset](#)
+
[Jupyter Notebook](#), [MBON Data Portal](#)

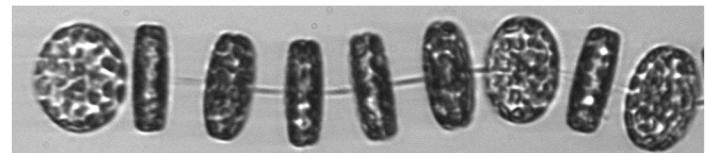
Species Observations



Taxon +
Location +
Date



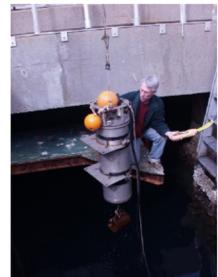
Imaging FlowCytobot (IFCB)



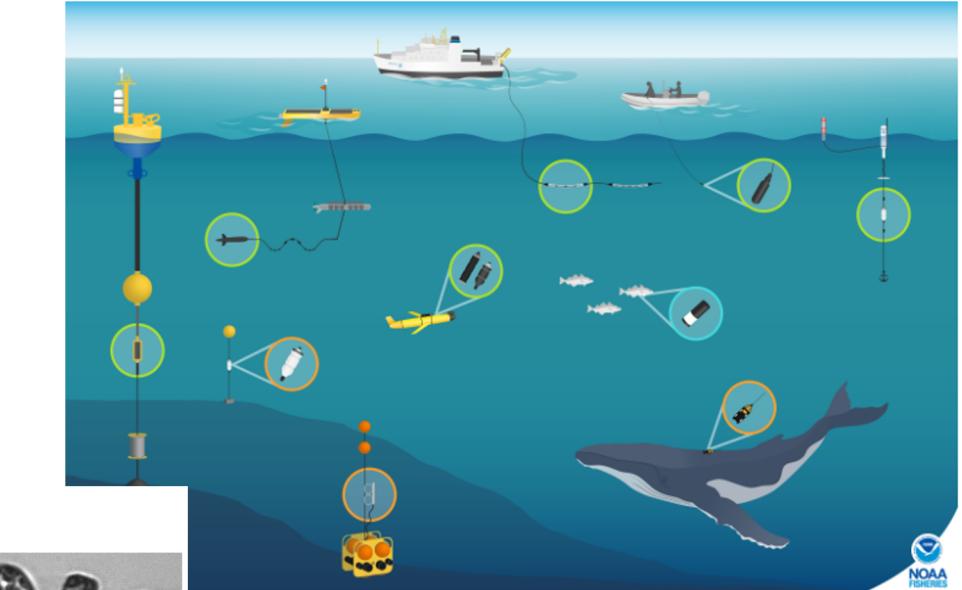
LONG TERM, HIGH-RESOLUTION MEASUREMENTS OF PHYTOPLANKTON ABUNDANCE AND CELL PROPERTIES: IMAGING FLOWCYTOBOT



Heidi with Imaging FlowCytobot in its pressure housing (R. Olson)



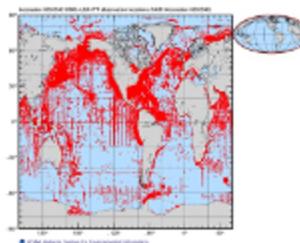
Rob Olson deploying Imaging FlowCytobot in the test well at the WHOI dock (A. Shalapyonok)



A fundamental understanding of the interaction between physical and biological factors that regulate plankton species composition requires, first of all, detailed and sustained observations. Only now is it becoming possible to acquire these types of observations, as we develop and deploy instruments that can continuously monitor individual organisms in the



Marine biological occurrence data managed by Ocean Biodiversity Information System USA (OBIS-USA) Integrated Publishing Toolkit (IPT) updated on 2022-09-28 (NCEI Accession 0250940)



Preview graphic

This dataset includes marine biological occurrence data managed by the Ocean Biodiversity Information System USA (OBIS-USA). The data were observed between 1832-11-30 and 2022-07-19. These data were last updated by OBIS-USA on 2022-09-28 and were received by NCEI on 2022-09-29.

OBIS-USA brings together marine biological occurrence data – recorded observations of identifiable marine species at a known time and place, collected primarily from U.S. waters or with U.S. funding. Coordinated by the Science Analytics and Synthesis (SAS) program of the United States Geological Survey (USGS), OBIS-USA, strives to meet national data integration and

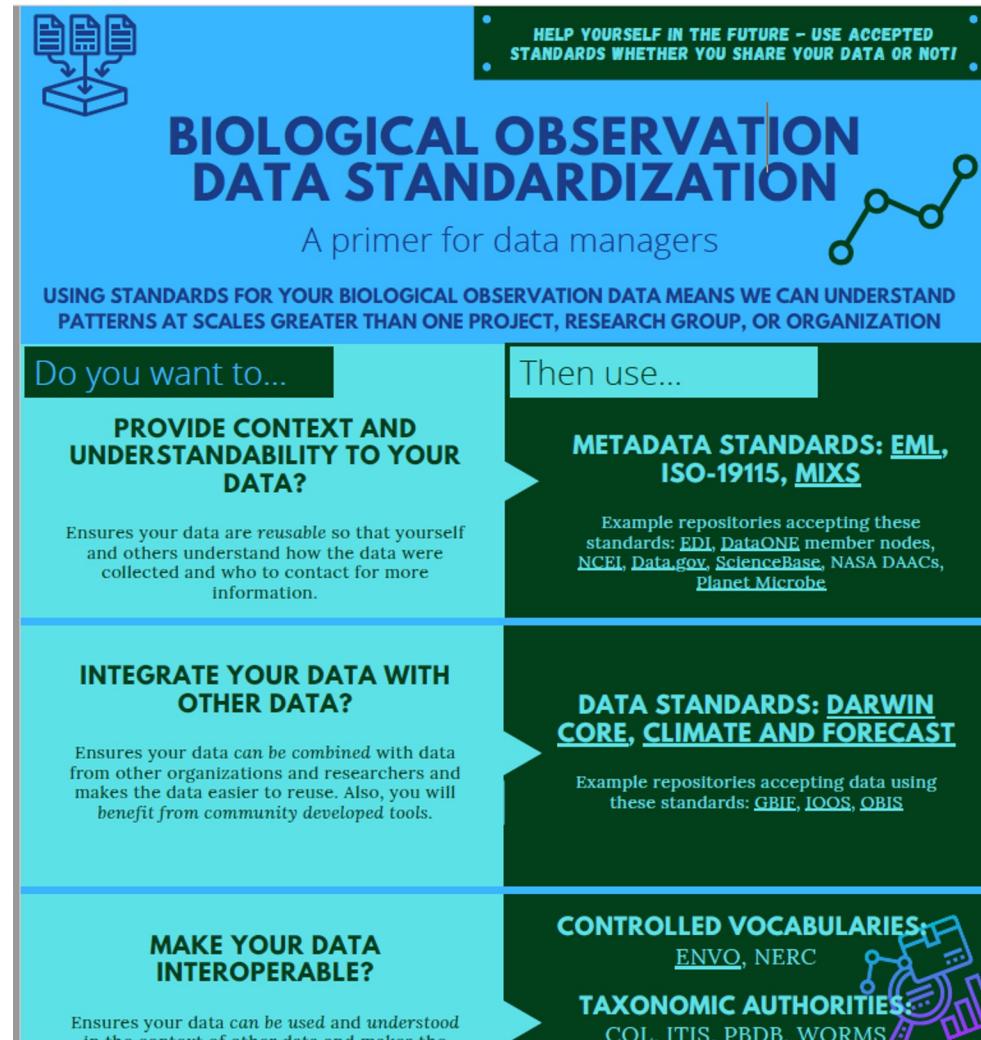
[Show more...](#)

[Dataset Citation](#)

[Dataset Identifiers](#)

[ISO 19115-2 Metadata](#)

Other updates



BIOLOGICAL OBSERVATION DATA STANDARDIZATION
A primer for data managers

USING STANDARDS FOR YOUR BIOLOGICAL OBSERVATION DATA MEANS WE CAN UNDERSTAND PATTERNS AT SCALES GREATER THAN ONE PROJECT, RESEARCH GROUP, OR ORGANIZATION

Do you want to...

PROVIDE CONTEXT AND UNDERSTANDABILITY TO YOUR DATA?
Ensures your data are reusable so that yourself and others understand how the data were collected and who to contact for more information.

Then use...

METADATA STANDARDS: EML, ISO-19115, MIXS
Example repositories accepting these standards: EDI, DataONE member nodes, NCEI, Data.gov, ScienceBase, NASA DAACs, Planet Microbe

INTEGRATE YOUR DATA WITH OTHER DATA?
Ensures your data can be combined with data from other organizations and researchers and makes the data easier to reuse. Also, you will benefit from community developed tools.

DATA STANDARDS: DARWIN CORE, CLIMATE AND FORECAST
Example repositories accepting data using these standards: GBIF, IOOS, OBIS

MAKE YOUR DATA INTEROPERABLE?
Ensures your data can be used and understood in the context of other data and makes the

CONTROLLED VOCABULARIES: ENVO, NERC
TAXONOMIC AUTHORITIES: COL, ITIS, PBDB, WORMS



<https://doi.org/10.6084/m9.figshare.16806712.v2>

MBON datasets in the pipeline

- Inventory form: <https://forms.gle/4NjKV7NWGLPcE2sb8>
- Inventory responses: <https://bit.ly/3WgyHTt>

What can we do with this information?

Metrics gathering:

https://github.com/MathewBiddle/sandbox/blob/main/notebooks/IOOS_occurrences_in_OBIS.ipynb

Global ES50 Indicator visualization (**in development**):
<https://mathewbiddle.github.io/globe/>

- Expectations for data management.
 - Blockers I anticipate for implementing the data management process are...
 - How much time/resources do you have allocated for data management activities?
 - What needs do you have regarding data management?

<https://app.sli.do/event/5p7hXw9ErjKm8CUucw6QcB>

Keep the conversation going

- Register datasets
- Assign data management specialist for each project.
- Next WG meeting Nov. 8th

MBON DMAC working group

Tuesday, November 8 · 1:00 – 1:50pm ET

Google Meet joining info

Video call link: <https://meet.google.com/imf-qrhy-jgx>

Or dial: (US) +1 321-351-6755 PIN: 548 540 777 #

More phone numbers: <https://tel.meet/imf-qrhy-jgx?pin=2704290817714>