

Hood Canal zooplankton data transformation, NANOOS / OBIS

Emilio Mayorga
Applied Physics Laboratory
University of Washington
emiliom@uw.edu
2023-4-25

IOOS & NANOOS context

- **<https://ioos.noaa.gov/>**
- **<https://ioos.noaa.gov/ioos-in-action/marine-life/>**
- **https://ioos.github.io/bio_data_guide/**
- Standardizing Marine Biological Data Working Group (SMBD)

OBIS

- What is OBIS? [**https://obis.org**](https://obis.org)
 - *Demo: Quick "Common name" search, using "Pacific krill". Then navigate from there, to map, stats, individual datasets, etc*
 - Example dataset: "Marine Mammals in Puget Sound" dataset (limited to individual presence):
[**https://obis.org/dataset/0e80dc63-b47c-423a-8e34-362f3171ea18**](https://obis.org/dataset/0e80dc63-b47c-423a-8e34-362f3171ea18) and [**https://mapper.obis.org/?datasetid=0e80dc63-b47c-423a-8e34-362f3171ea18#**](https://mapper.obis.org/?datasetid=0e80dc63-b47c-423a-8e34-362f3171ea18#)
 - Recent team project in OceanHackWeek in Spanish event: [**GitHub repository**](#) and [**Presentation pdf**](#)

- Greater reach and access to the data: [**https://obis.org/library/**](https://obis.org/library/)
- Feeding data into the Marine Biodiversity Observation Network (MBON), [**https://marinebon.org**](https://marinebon.org)
- OBIS US Node: Abby Benson, USGS
- Data standards requirement. **Darwin Core**, including "extended measurements or facts" eMoF extension.
[**https://manual.obis.org**](https://manual.obis.org). Makes data and metadata more unambiguous, interoperable and "machine-friendly"

Why the Hood Canal dataset

- Dataset previously submitted to and available at BCO-DMO
 - BCO-DMO Dataset 682074: **Zooplankton densities collected from a seasonally hypoxic fjord on R/V Clifford A Barnes cruises from 2012-2013 (Pelagic Hypoxia project)**. ERDDAP dataset: **https://erddap.bco-dmo.org/erddap/tabledap/bcodmo_dataset_682074.html**. A description of each data column is found in the "Parameters" section in **<https://www.bco-dmo.org/dataset/682074>**
 - BCO-DMO Project 557504: **Consequences of hypoxia on food web linkages in a pelagic marine ecosystem (PelagicHypoxia)**

- Focus on an already published and well-organized dataset allowed me to focus this first NANOOS dataset transformation project on the "Darwin Core data alignment" and OBIS submission tasks.
- Highlight: Advantages / extra capabilities of data in OBIS vs dataset downloaded from BCO-DMO

What I've done (high-level summary)

- See <https://github.com/nanoos-pnw/obis-keisterhczoop> (I'll update this by next week)
- Downloaded data from BCO-DMO (single csv, supplemented with metadata)
- Explored the dataset and familiarized myself with it. Included assessing consistency and reading a paper from the data (Li et al, 2019)
- Developed code (Python in Jupyter notebook) to apply transformations (below) to Darwin Core standard
- Worked with OBIS (Abby Benson) and the SMBD group to get input on specific strategies, to review initial draft of data alignment, and to get initial overview of the process for

submitting the transformed data once finalized

Transformations ("data alignment")

Types of transformations:

- Reorganize into required tables, with distinct, interrelated information: events, occurrences, and "measurements or facts"
- Use required column names
- Transform values to expected formats (eg, dates and times)
- Transform certain types of entries to expected, universal vocabularies. eg, taxonomy mapped to World Register of Marine Species (WoRMS), <https://www.marinespecies.org>; life stage (incomplete matches), <http://vocab.nerc.ac.uk/collection/S11/current/>
- Enrich information through more explicit (and usually more

verbose) presentations (eg, "cruise" and "station visit" events)

- Error checking
 - The questions I asked about missing times, timezones, life stages

Transformations, continued

Walk through some parts of each notebook and "mapping" tables/JSON, then show each resulting table. Provide examples of the transformations.

Next steps

- Review (Spot check? Ask questions?): Amanda and BethElLee
- Finalize
- Submit!
- Assess feasibility of doing the same for **Puget Sound zooplankton dataset**