

Standardizing Marine Biological Data Working Group update

Tim van der Stap, MSc.
Scientific Data Specialist
Hakai Institute

IOOS DMAC, September 2023



Background

Conceptualized at IOOS Code Sprint in Ann Arbor 2019

*"The purpose of the SMBD is to facilitate a community of practice for **aligning marine biological data** to **Darwin Core** for sharing to **OBIS**. We do this by empowering our community members with the tools and knowledge to mobilize marine biological data."*

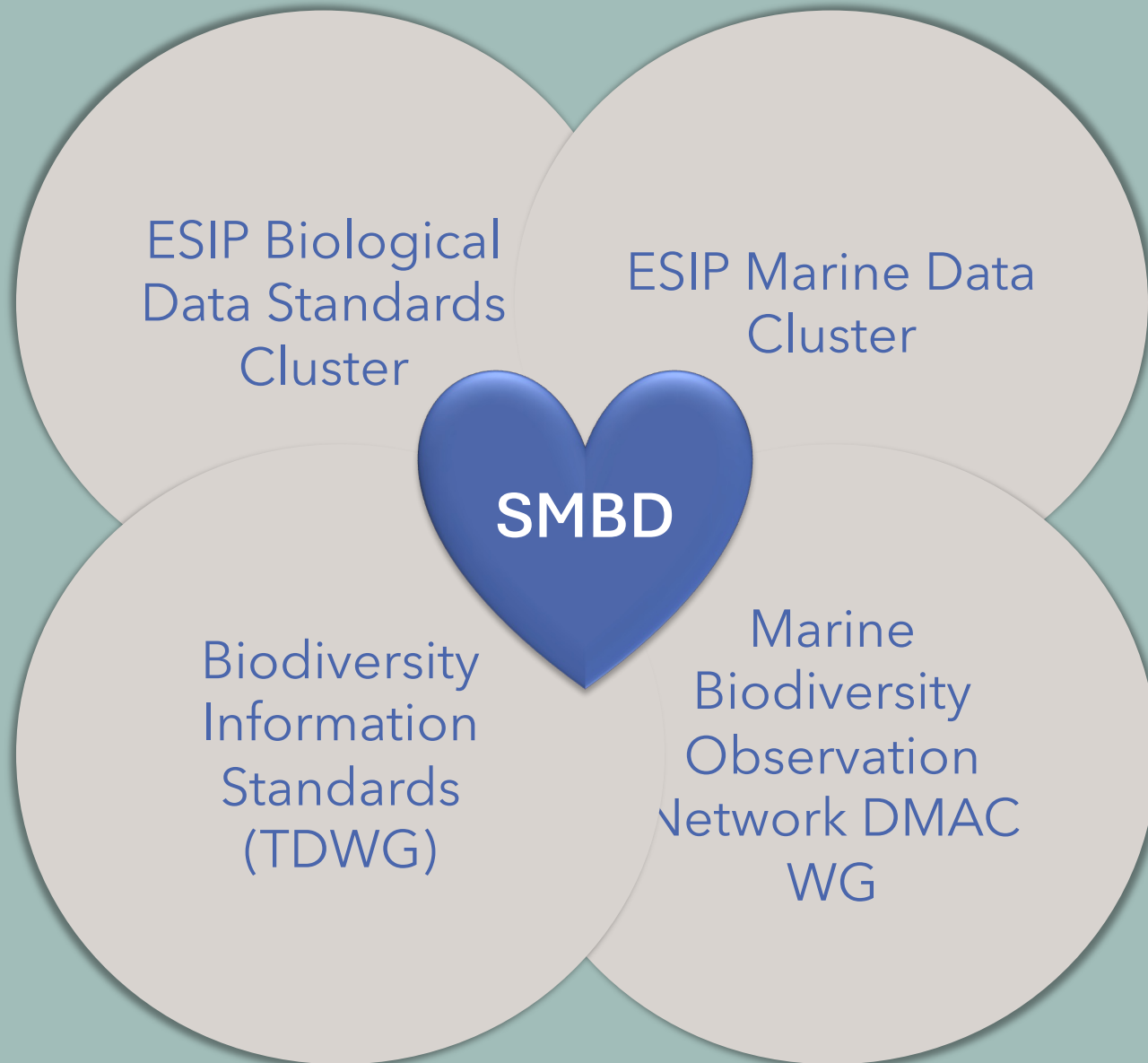


The background of the slide features a collage of nature photographs arranged in a honeycomb pattern of hexagonal frames. The photos include: a sandy beach with a large, dark, rocky formation in the ocean; a close-up of a giraffe's head and neck showing its characteristic patterned skin; a close-up of a turtle's head and shell; and a close-up of a bird's head. The overall color palette is muted, with a light teal background for the text area.

Monthly Meetings

- Practical and informal in nature
- Participants from various marine institutions
- Discussing relevant topics, blockers, updates and data questions
- GitHub repository:
https://github.com/ioos/bio_data_guide
- Slack: 215 members - data questions, blockers, announcements

Contextualizing SMBD



- Marine biological data
- Darwin Core / OBIS and associated standards
- WoRMS, NERC Vocabulary Server
- **Practical** by nature



SMBD GitHub Repository

- **Datasets** actively being worked on by community members
- **Code** and **documentation** used on other datasets that can be re-used
- **Tools** to help you navigate the organizational or technical challenges of publishing to OBIS.
- Link to use cases / examples in the **Darwin Core Marine Example Compendium** - a living 'cookbook'



https://ioos.github.io/bio_data_guide/index.html



Annual Marine Biological Data Mobilization Workshop

- Held in 2022 and 2023:
<https://zenodo.org/record/7896606>
- Run by members of the SMBD Working Group
- Reach beyond the SMBD Working Group
- **International** in scope: several dozen scientists from 4 continents, 17 countries
- Helped **mobilize** 2 datasets to OBIS, +/- 50 datasets in progress

Marine Biological Data Mobilization Workshop 2023

Introduction to Darwin Core

Overview

Teaching: 0 min
Exercises: 90 min

Questions

- What is Darwin Core?
- What is a Darwin Core Archive?
- Why do people use Darwin Core for their data?
- What are the required Darwin Core terms for sharing to OBIS?

Objectives

- Understand the purpose of Darwin Core.
- Understand how to map data to Darwin Core.
- Plan for mapping to Darwin Core.

Darwin Core - A global community of data sharing and integration

Darwin Core is a data standard to mobilize and share biodiversity data. Over the years, the Darwin Core standard has expanded to enable exchange and sharing of diverse types of biological observations from citizen scientists, ecological monitoring, eDNA, animal telemetry, taxonomic treatments, and many others. Darwin Core is applicable to any observation of an organism (scientific name, OTU, or other methods of defining a species) at a particular place and time. In Darwin Core this is an **occurrence**. To learn more about the foundations of Darwin Core read [Wieczorek et al. 2012](#).

Demonstrated Use of Darwin Core

The power of Darwin Core is most evident in the data aggregators that harvest data using that standard. The one we will refer to most frequently in this workshop is the [Ocean Biodiversity Information System](#) (learn more about OBIS). Another prominent one is the [Global Biodiversity Information Facility](#) (learn more about GBIF). It's also used by the Atlas of Living Australia, iDigBio, among others.

Darwin Core Archives

Darwin Core Archives are what OBIS and GBIF harvest into their systems. Fortunately the software created and maintained by GBIF, the [Integrated Publishing Toolkit](#), produces Darwin Core Archives for us. Darwin Core Archives are pretty simple. It's a zipped folder containing the data (one or several files depending on how many extensions you use), an Ecological Metadata Language (EML) XML file, and a meta.xml file that describes what's in the zipped folder.

Exercise

Challenge: Download this [Darwin Core Archive](#) and examine what's in it. Did you find anything unusual or that you don't understand what it is?

Solution

Darwin Core Mapping

Now that we understand a bit more about why Darwin Core was created and how it is used today we can begin the work of mapping data to the standard. The key resource when mapping data to Darwin Core is the [Darwin Core Quick Reference Guide](#). This document provides an easy-to-read reference of the currently recommended terms for the Darwin Core standard. There are a lot of terms there and you won't use them all for every dataset (or even use them all on any dataset) but as you apply the standard to more datasets you'll become more familiar with the terms.

Tip

If your raw column headers are Darwin Core terms verbatim then you can skip this step! Next time you plan data collection use the standard DwC term headers!

https://ioos.github.io/bio_mobilization_workshop/

Features

- Presentations on actual, current-day relevant topics
 - eDNA: DNA Derived Data Extension
 - Passive Acoustic Monitoring (PAM) data
 - Let us know if there are any topics **you** would like to see discussed!

Moving forward

- Community free to join
 - **Meetings:** every 2nd Wednesday of the month at 16:00 ET
 - **Slack Channel**
 - **SMBD GitHub Repository**
- 2024 Marine Biological Data Mobilization Workshop
 - Date TBD!
- Questions, comments or feedback
 - tim.vanderstap@hakai.org



[Standardizing Marine
Biological Data Slack Channel](#)



👉 Join us every month: scan QR code for connection details 👉