BIOLOGICAL OBSERVATION DATA STANDARDIZATION

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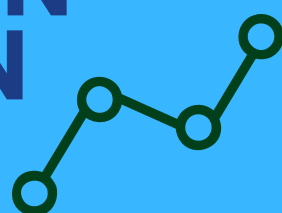
USING STANDARDS FOR YOUR BIOLOGICAL OBSERVATION DATA MEANS WE CAN UNDERSTAND PATTERNS AT SCALES GREATER THAN ONE PROJECT, RESEARCH GROUP, OR ORGANIZATION

Do you want to...

Then use...

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.

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 data easier to reuse, especially for machine-to-machine operations.

CONTROLLED VOCABULARIES: ENVO, NERC

TAXONOMIC AUTHORITIES: COL, ITIS, PBDB, WORMS

HABITAT CLASSIFICATION: CMECS, NVCS, NWCS

MAKE YOUR DATA INTERNET READY?

Ensures your data are *programmatically accessible* so it can be shared online.

WEB-ENABLED STANDARDS: DUBLIN CORE, DATACITE, SCHEMA.ORG

Example places accepting these standards: Zenodo, Pangaea, FigShare, Dryad, University or Organization system

EXAMPLE WEB SERVICES: ERDDAP, MAP SERVICES, THREDDS

MAKE YOUR DATA SOFTWARE READY?

Ensures your data are *easily imported to computing software* and analyzed.

- Use non-proprietary formats
- Structure data in tidy/long format
- Follow ISO 8601 for dates
- Match scientific names to a taxonomic authority
- Record latitude and longitude in decimal degrees in WGS84
- Use globally unique identifiers

MIX AND MATCH BUT WATCH OUT FOR REQUIREMENTS!

You can use metadata standards, data standards, and controlled vocabularies in any combination with each other. Some places may require certain combinations for your data to be included in their repository. Using standards will result in data that are more closely aligned to the FAIR principles.

REQUIRED STANDARDS FOR SOME EXAMPLE REPOSITORIES

EDI: EML
OBIS: Darwin Core, EML, and WoRMS, NERC recommended
GBIF: Darwin Core and EML
NCEI: ISO
Planet Microbe: MIxS and ENVO

Visit **https://wiki.esipfed.org/Biological_Data_Standards_Cluster** to get involved

https://fairsharing.org/standards/?q=biological