Data Overview

Holly Little ✦ littleh@si.edu
with contributions from GBIF

Smithsonian Institution

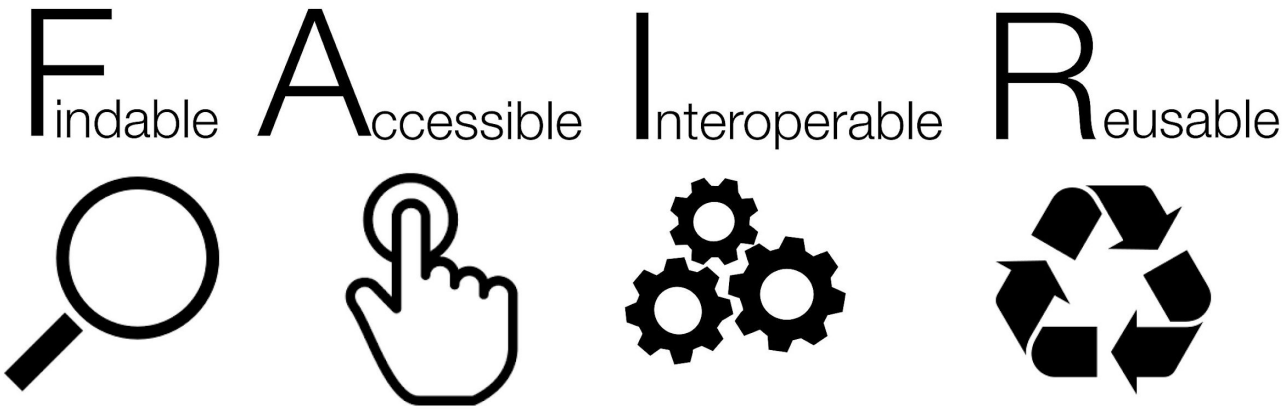
NATIONAL MUSEUM of NATURAL HISTORY

GBIF
Global Biodiversity Information Facility

biodiversity_next

better data - better science - better policies

“From local data to global impact, but also from global data to local solutions.”

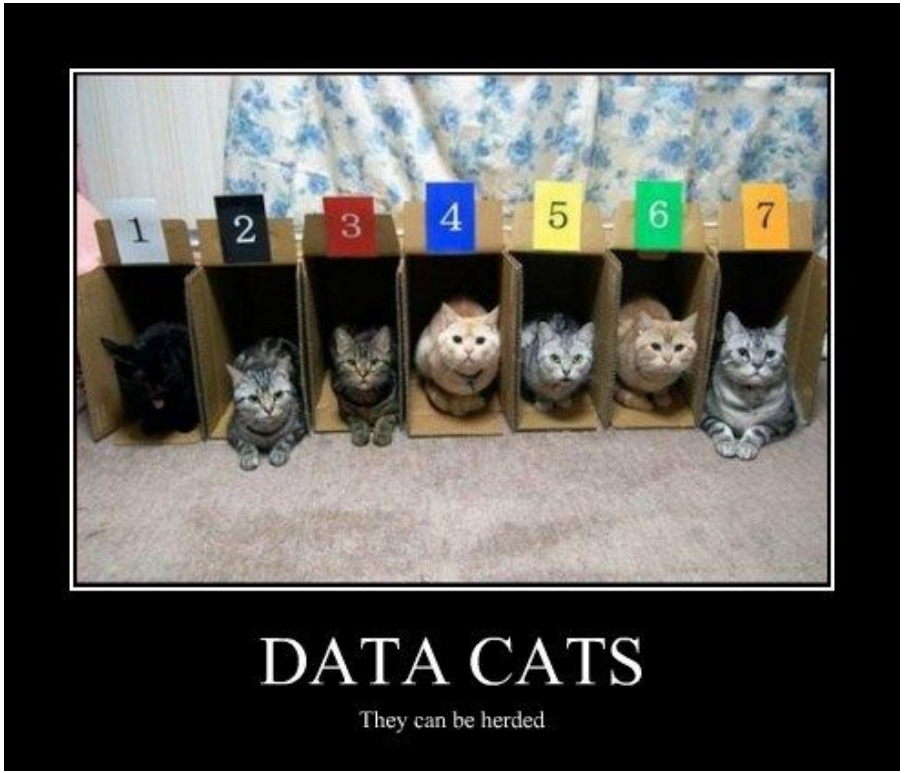


https://en.wikipedia.org/wiki/FAIR_data

How do you classify data?

Many different ways:

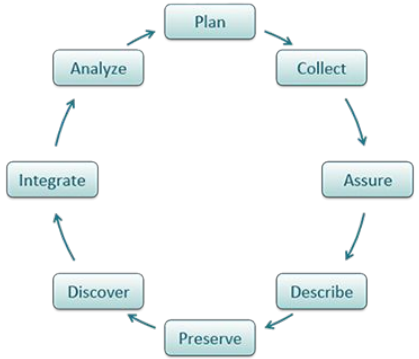
- By discipline?
- By source?
- By use?
- By function?
- By track?
- By method?
- By format?



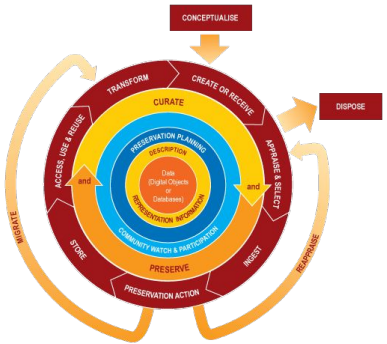
cheezeburger.com

Some considerations for all data - Data Lifecycle

- How is the data collected?
- Interoperability: What standards are needed? What are the best practices for implementation?
- How can we expand on the data, increasing quality and usability?
- How well is the data documented?
- How is the data managed? Both short and long term.
- What data governance is needed?
- How is the data disseminated? Are there restrictions?



https://www.dataone.org/data-life-cycle



http://www.dcc.ac.uk/resources/curation-lifecycle-model

Data Type Outlines

Data Type

Explanation

- *Example text*

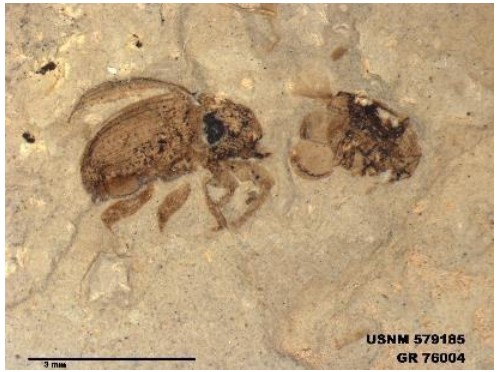
Example image

biodiversity_next pointers

Occurrence Data: Collected Specimens

Occurrence data consists of individual organisms in time and space. Collected specimens can be:

- Preserved Specimens
- Fossil Specimens
- Living Specimens
- Geologic Samples



WS57: Promoting natural history collections for understanding biodiversity and biodiversity change

WT44: TDWG Paleo IG Workshop: Coordinating best practices for fossil specimen data mobilization

Occurrence Data: Observations

Occurrence data is also generated through observations of an organism:

- Human Observations
- Machine Observations (also Sensor Data)

https://emammal.si.edu

iNaturalist https://www.inaturalist.org/

SI37: Making long term ecological research data FAIR

ST92: Advancing the Quality of Diverse Citizen Science Data from Observations to Atlasing Projects

ST69: Machine Observations Interest Group: A first dive into Darwin Core for exchanging biologging data

Taxonomic Data

Information about the earth's species and the evolutionary relationships among them. Summary or baseline inventory of taxa in a given context, such as:

- Checklists
- Taxon Reference Lists
- Authorities
- Classification Schemes
- Names

Courtesy GBIF

ITIS

https://www.itis.gov/

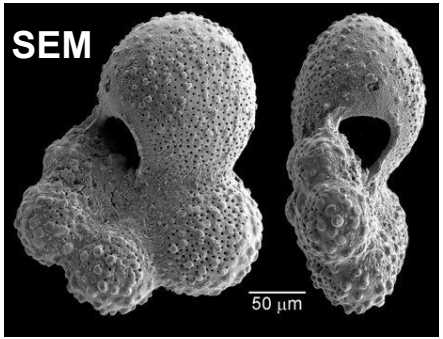
ST13: Enhancing taxonomic publications for dynamic data exchange and navigation

SI87: Empowering the taxonomic community by linking information through names and taxonomy

Media Data

Digital surrogates of biodiversity:

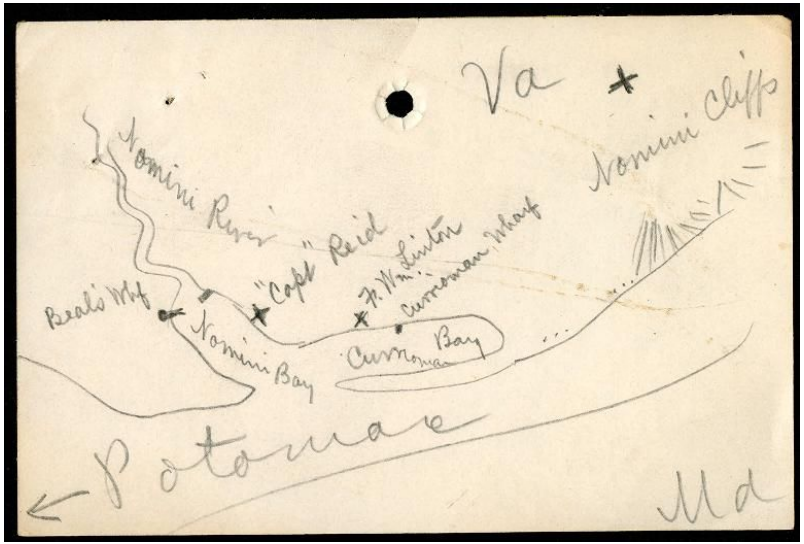
- 3D
- 2D
- CT
- Audio
- Video



Literature and Archival Data

Metadata describing published or archival materials and metadata extracted from the content of these materials often contributing to other data types:

- Publications
- Reports
- Field notes and logs
- Maps
- Protocol Notes
- Surveys and assessments



biodiversitylibrary.org/item/181019

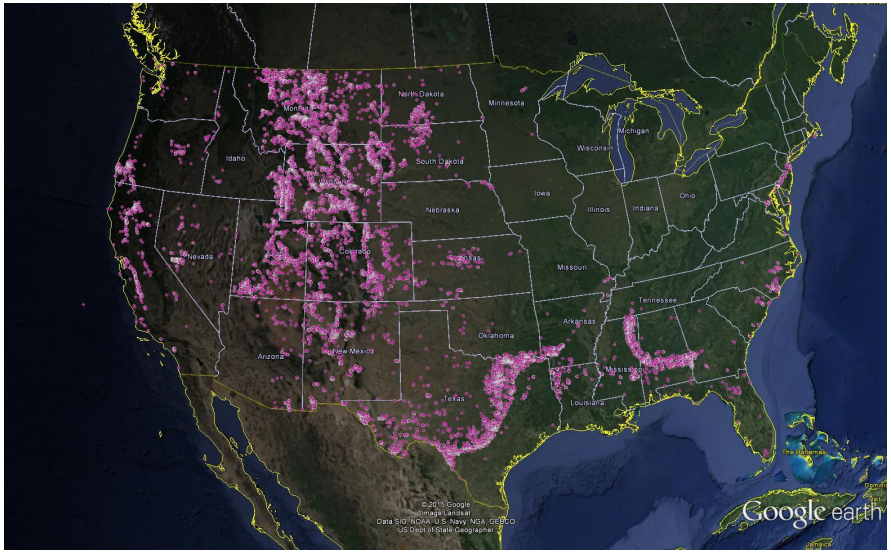
SI33: Improving access to hidden scientific data in the Biodiversity Heritage Library

WI50: Accessing knowledge from legacy biodiversity literature

Sensor Data

A type of machine observation data, gathered by device:

- GPS
- Radar or Satellite Data (also Occurrence Data)
- Camera Trap (also Imaging Data, also Occurrence Data)
- Seismic Stations



USGS Localities provided by Casey McKinney, USGS

SS31: Quantification of biodiversity across scales

Genomic Data

The genome and DNA data of an organism:

- DNA barcoding
- FASTA files
- e-DNA

```
>BAH8001_Coryphaena_hippurus
NNNNNNNNNNNNNNTCGGTGTCTTAGCAGGGATAACAGGAACAGGTTTAAGTCTTCTCATTCGAGCTGAG
>BAH8002_Coryphaena_hippurus
CCTTTATTTAATTTTCGGTGTCTTAGCAGGGATAACAGGAACAGGTTTAAGTCTTCTCATTCGAGCTGAG
>BAH8003_Coryphaena_hippurus
NNNNNNNNNNNNNNTCGGTGTCTTAGCAGGGATAACAGGAACAGGTTTAAGTCTTCTCATTCGAGCTGAG
>BAH8004_Ocyurus_chrysurus
CCTTTATCTAGTATTTGGTGCCTGGGCCGGAATAGTAGGCACGGCCCTAAGCCTGCTCATTCGAGCAGAA
>BAH8005_Ocyurus_chrysurus
CCTTTATCTAGTATTTGGTGCCTGGGCCGGAATAGTAGGCACGGCCCTAAGCCTGCTCATTCGAGCAGAA
>BAH8006_Ocyurus_chrysurus
CCTTTATCTAGTATTTGGTGCCTGGGCCGGAATAGTAGGCACGGCCCTAAGCCTGCTCATTCGAGCAGAA
>BAH8007_Scomberomorus_regalis
CCTCTATCTAGTATTTGGTGCATGAGCTGGAATAGTTGGCACAGCCCTAAGCCTTCTTATCCGAGCTGAA
>BAH8008_Epinephelus
NNNNNNNNNGTATTTGGTGCCTGAGCTGGTATAGTAGGAACAGCCCTAAGCCTGCTCATTCGAGCTGAG
>BAH8009_Epinephelus
CCTCTATCTTGTATTTGGTGCCTGAGCTGGTATAGTAGGAACAGCCCTAAGCCTGCTCATTCGAGCTGAG
```

ST15: Molecular biodiversity evidence in time and space: data linkages and standards

SI29: GGBN – A global infrastructure for molecular research and collections

Data types listed so far...

- Occurrence Data
- Taxonomic Data
- Media Data
- Literature and Archival Data
- Sensor Data
- Genomic Data

In progress...

- *Sampling Event Data*
- *Trait Data*
- *Interaction Data*
- *Isotopic Data*

https://bit.ly/bi101data

Data and You

What kind of data do you collect? Or What kind of data do you work with?

Questions?

Holly - littleh@si.edu

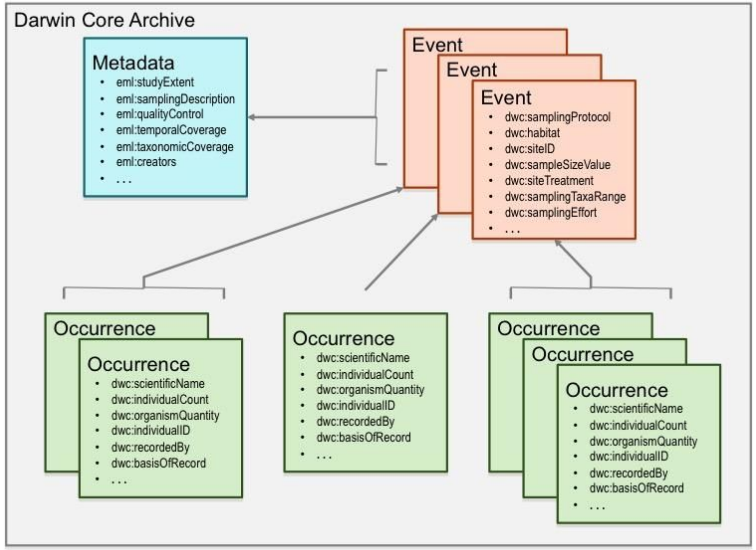
Acknowledgements

Unless otherwise noted, images courtesy of the Smithsonian National Museum of Natural History (@nmnh)

Sampling Event Data

Lists of collecting events and their observed species, together with data on sampling methods and often with quantitative information:

- Systematic Monitoring Schemes



https://www.gbif.org/sampling-event-data

SP35: Operationalizing Essential Biodiversity Variables: data integration, production and dissemination

Trait Data

Information about the qualities of an organism:

- Morphometric Data

Example

SI72: Operationalizing Trait-Based Biodiversity

Interaction Data

Explanation

Examples

biodiversity_next pointers

Isotopic Data

- Isotope Ecology and Paleoecology

Examples

biodiversity_next pointers