Open Geoscience Data:
the Role and Value of [Domain] Repositories

Credit: Pixabay

Danie Kinkade

BCO-DMO Team: Adam Shepherd, Mak Saito, Peter Wiebe, Shannon Rauch, Karen Soenen, Amber York, Dana Gerlach, Sawyer Newman, Taylor Heyl

Woods Hole Oceanographic Institution

April, 19, 2023

BCO-DMO
Biological & Chemical Oceanography Data Management Office

Award: 1924618

Repository Contribution to the Open Science Enterprise

- Partner with investigators to
  - Educate on benefits of openly sharing research output
  - **Facilitate FAIR data sharing**
- Assist in stewardship of entire research effort
  - Direct researchers to appropriate repositories for sharing
  - Linking distributed data to create full picture for reuse
- Engage broader research data publishing community

Helping Raise Awareness: Why Share Research Data?

Benefits to...

Researcher:

- Are credited for your hard work!
- Can increase citations
- Can boost collaborations
- Can increases exposure
- Satisfy funder requirements

Research Community:

- Builds a reusable resource
- Facilitates new discoveries
- Sparks new collaborations
- Enables transparency of research results

Society:

- Transparency boosts public confidence in scientific process
- Contribute to management and policy
- Access by non-research audiences (education, general public)

Photo credits: WHOI image bank

Sharing is Hard: Researchers Can't Go It Alone...

Findable: Data are linked to descriptive persistent metadata.

Accessible: Data and metadata are open, free, and machine accessible.

Interoperable: Data and metadata are standardized and use vocabularies. Data points to related metadata.

Reusable: Metadata are rich, and employ usage licenses, provenance, and community standards.

Wilkinson, M.D., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. , https://doi.org/10.1038/sdata.2016.18

Oceanographic Data are Heterogeneous and Complex!

Disciplines:

- Biological
- Chemical
- Biogeochemical
- Physical
- Geophysical

Collection Types:

- *In situ*
- Laboratory
- Remotely sensed
- Synthetic/derived

Additional Challenges:

- Variable organization
- Varying metadata
- Local parameter terms
- Emerging data types
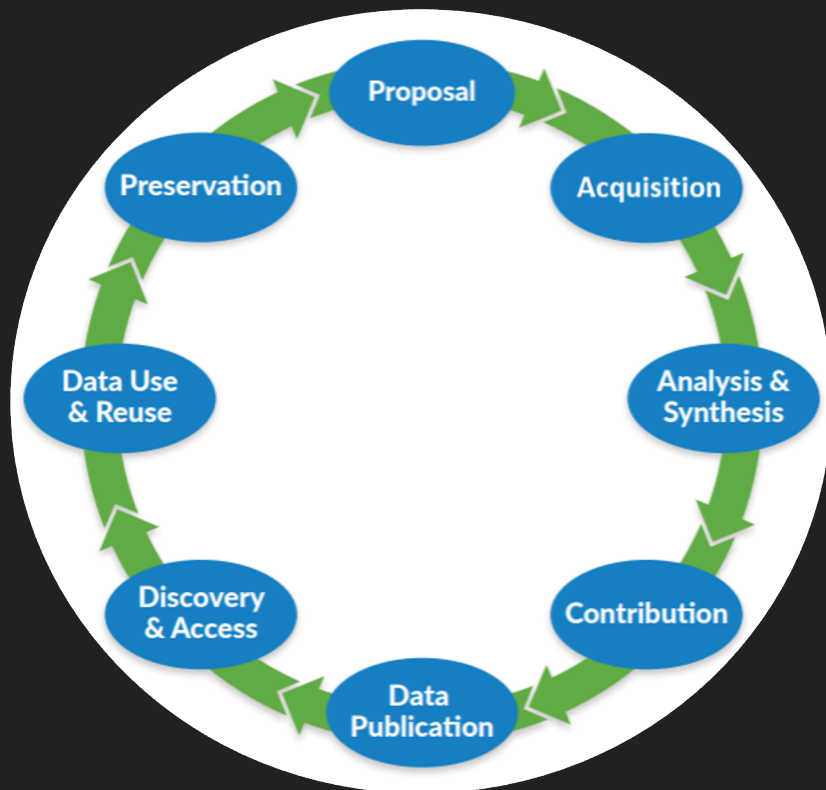- Distributed complementary info

Formats:

- ASCII Text (tabular)
- Binary (e.g., NetCDF)
- Images
- Acoustics
- Application (e.g., Matlab)
- Links to other data

Scale:

- Molecular to Megafaunal
- Local to Global
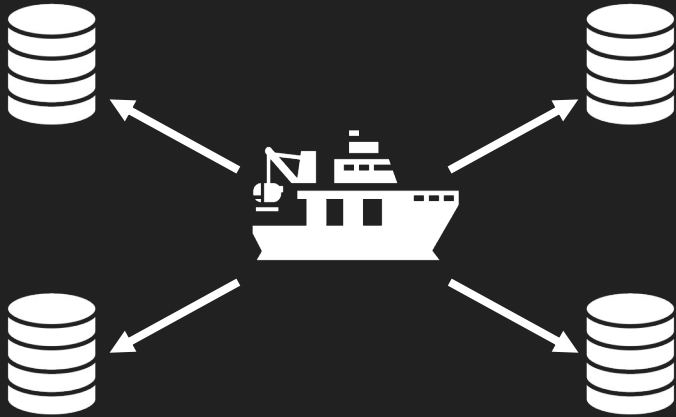- Discrete to continuous /synoptic

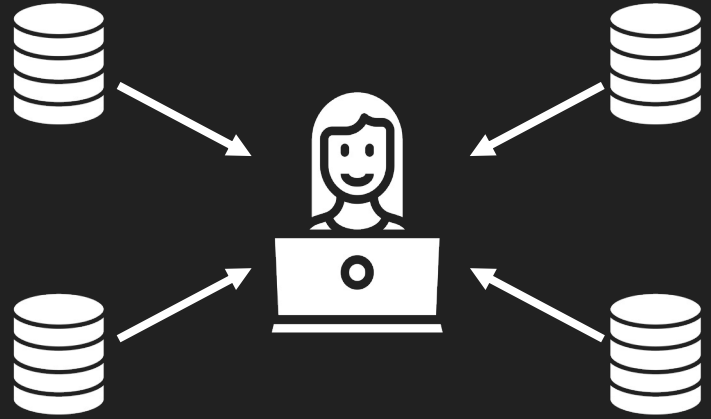# Supporting Data Curation and Sharing Throughout its Lifecycle

- Guidance on data management planning, data formatting and standards;

- Improve data quality and interoperability;

- Capture and record rich metadata;

- Publish with licenses, digital object identifiers (DOIs) and recommended citations.

- Provide access to data and metadata (restricting access as appropriate);

- Track usage metrics

- Ensure final, long-term archive

Enabling Connections to Create a Full Picture of the Research Process

- Research output may need curation by multiple repositories

- Linking distributed data and information sources seamlessly for discovery and interoperability

Engaging with Broader Data Publishing Ecosystem

REPOSITORIES

Domain repositories relieve researchers from the difficult aspects of data sharing, educate them in better data management practices that will improve the process of conducting open and transparent research and contribute a rich resource of data and information to fuel the scientific endeavor.

(Lawson, 2002)

Notes from Shelley

domain repositories contribute to the Open Science enterprise by…

Partnering with researchers to educate them on benefits of openly sharing their data, but also facilitate FAIR data sharing

> Why share?

> Researchers may be able to make data accessible, but only repos can make data FAIR

> Highly heterogeneous data - need special curation approaches

> Repos can help tie the research picture all together:

> > Help researcher find appropriate repositories for their many data products

> > Help users see big picture of research

Repos relieve the researcher's burden of the challenging aspects of data sharing