Intersection of the Data Repository and Journal Workflows:

Coordination of Data Preservation with the Paper Peer-review Process

Session Chairs:
Danie Kinkade (BCO-DMO); **Shelley Stall** (AGU); **Lesley Wyborn (**ANU)

12 April 2021
4:00pm - 5:30pm EDT

Workshop Co-Chairs

Danie Kinkade- Director, Biological and Chemical Oceanography Data Management Office (BCO-DMO)

Shelley Stall- Sr. Director for Data Leadership, American Geophysical Union (AGU)

Lesley Wyborn- Honorary Professor, National Computational Infrastructure Facility and Research School of Earth Sciences, Australian National University (ANU)

Agenda

- Introduction to the problem
  - Why do we need to come together and how can a workflow diagram help?
  - History of work to date
- Two brief talks: repository and journal perspectives
- Review of journal workflow and needs
- Open discussion
- Consider draft recommendations
- Next steps

Why Do We Need a Workflow Diagram?

- Domain and institutional repositories are challenged to fit into publication workflows

- Bottlenecks are arising in touchpoints between stakeholder workflows

- Little awareness and communication across relevant stakeholders

- Need consensus and common frame of reference to understand issues

- Serves as framework to collaboratively address problems and improve efficiency of data flow from researcher to repository to publisher

- Sprang from conversations within Council of Data Facilities Webinars in Feb 2020

- Concept socialized at ESIP Summer Meeting 2020 COPDESS session

- Presented at RDA VP16 in the ESES IG Session Breakout on repositories

  - Recognition of RDA/WDS Publishing Data Workflows WG
    (Dallmeier-Tiessen, et al., 2016 http://doi.org/10.15497/RDA00004)

- Refinement and continued sharing at AGU FM 2020 and ESIP Winter 2021

- Further refinement now between Repositories & Journals - hope to share outcomes/recommendations among all stakeholders

Considerations to keep in mind...

- Focus is on geoscience observational data (vs. samples or code)

- Focus is on new data submissions vs. reuse

- Long tail vs operational/streaming

- Probably incomplete!

- **Need to zero in repository-journal touch points!**

http://bit.ly/dataPubWkflo
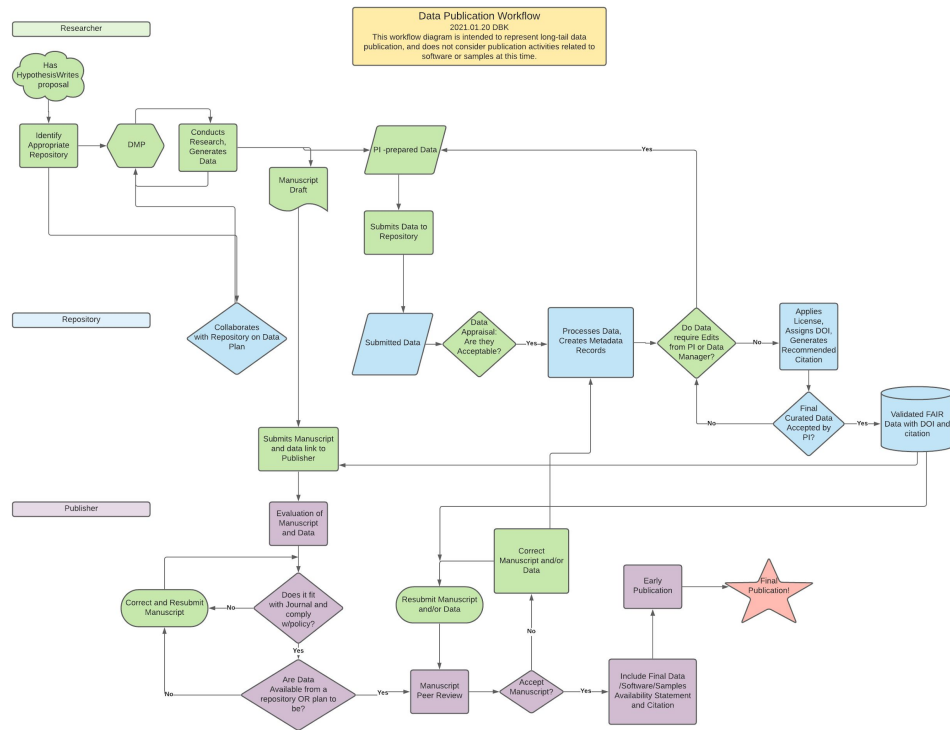
Workshop Goals

- Transparency of repository workflow and the touchpoints with the author in regard to the data.

- Transparency of the journal workflow concerning data availability and citation and when it is needed.

- Draft recommendations (or at least a sketch)
  - For journals on how to better support repositories to make this workflow easier.
  - For repositories on how to better support journals.

The Rest of This Workshop...

- Provide context on the example use case

  - Repository perspective

    - BCO-DMO (Domain, long-tail)

  - Publisher perspective

    - AGU

- Identify proactive actions both journals and repositories can take to have a smoother workflow

- Consider implementation recommendations for presentation at RDA PV17

BCO-DMO Data Curation and Publication Workflow
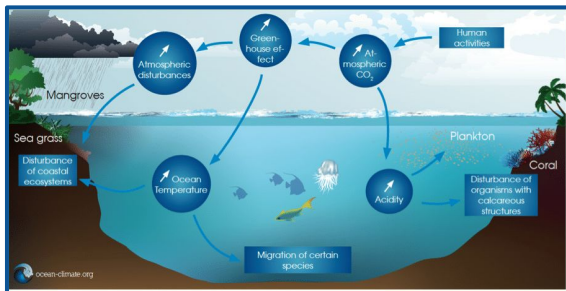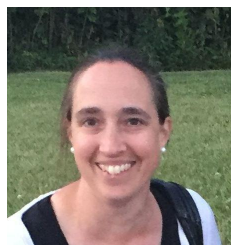
BCO-DMO Data Curation and Publication Workflow
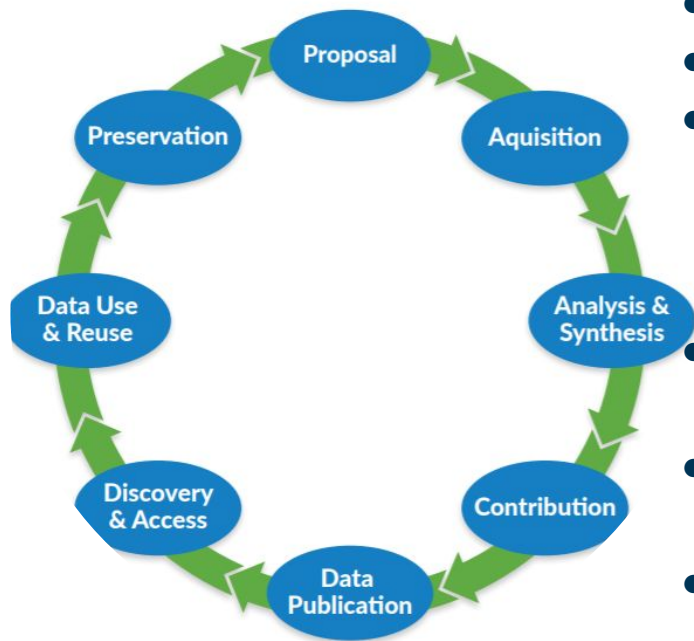
- Engages researchers early in data management planning.
- Work with each investigator to process and publish data.
- Applies domain knowledge to increase interoperability: harmonizes formats, and observational precision and metadata; performs gross QC; creates domain-appropriate metadata.
- Applies licensing, DOIs, creates citations, and documents provenance and versioning.
- Employs emerging technologies, community best practices and standards to improve discoverability, access, and reuse.
- Educates our research community on domain data best practices.

BCO-DMO Data Curation and Publication Workflow

Pain Point!

9 January 2021

Hi BCO-DMO — ughh. The proofs have arrived this week, much to my surprise. This does not seem feasible to add the DOIs to the paper, unless you advise otherwise. I am sorry this is so hard to tie the threads together.

10 January 2021

… no worries. It is not a reasonable request.

What would happen if I requested data doi when I submit the paper and then deal with the risk that the paper might get rejected?

How Do We...

- Encourage (researchers, funders) to

  - Submit full project output

  - Submit data once ready for analysis (i.e., as early as possible!) not at time of scholarly publication

    - How to better align what's needed for publication vs full project output?

- Manage needs of publishers while providing robust curation of data?

Now from the journal perspective...

Use Case: The paper is being submitted to a journal before the supporting data is fully preserved.

Manuscript submission (author) → Quality Check (staff) → Editor Review → Peer Review → Peer Review Response (editor) → Manuscript Accepted

Availability Statement:

Author indicates in an **intentional, temporary statement** of "what data" and "where" is it to be preserved.

Availability Statement:

The state of the statement is assessed. Guidance to author is provided if intentional statement is missing.

Availability Statement:

Should the editor want to move forward with the paper, the **data now needs to be available for peer review** and support evaluation of the paper. This could be a PID, share link, or the data is provided temporarily in other ways.

Availability Statement:

The availability statement is updated with the **final information** about the data and PID once it is ready from the repository.

Availability Statement:

The statement must be **compliant**.

Citation:

Not yet available

Citation:

Not yet available

Citation:

Not yet available

Citation:

Added to the Reference Section.

Citation:

Citation is present and compliant with format.

The Basic Steps...

- Author guidelines

- Paper submitted (author)

- Quality check of paper (staff)

- Editor review

- Peer review

- Final manuscript decision -- acceptance

- Journals can...
    - Help authors **select the appropriate** domain/institutional/ general repository for the data (if they have not done so yet).

    - Give the author a **sense of timing** that the repository needs to curate and preserve the data.

    - Information on the repositories **use of share links for use during paper peer review** prior to paper publication.

    - Details about when the repository makes the data public, the paper being accepted, the DOI for the data, and timing in general specific to the repository process.

    - What else?

How can we smooth this workflow…

- Repositories can…

  - The ability for the paper peer reviewer to confidentially access the data specific to the paper. This can be done by a share link, or other means.

    - Some journals have double-blind peer review (T&F) and can only use limited number of repositories that support that.

  - The assurance and commitment from the repository that the peer reviewer identity will be kept confidential. A statement pertaining to this on the repository page is helpful here.