The Intersection of the Data Repository and Journal Workflow -- Coordination of Data Preservation with the Paper Peer-Review Process

Virtual Workshop

April 12, 2021

1600 - 1730 EDT

Workshop 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)

[PLEASE SIGN IN](#6iixtbkdp0c7) ⇐ Please go here!!

Workshop materials:

* [Data Publication Workflow: Publisher Side](https://drive.google.com/file/d/1d0-xUhp_rUOJ01SaQhamepPrdZiuUoRp/view?usp=sharing)
* [Workshop Presentation Slides](https://docs.google.com/presentation/d/18d3GW3y462RakCR3Uudu-L7Rw-LOhxolP0FSWxiImvA/edit?usp=sharing)

Driving Problem:

Increasing effort is being directed toward streamlining the activities surrounding research data publication, especially with respect to data supporting the findings of scholarly publications. Yet, we still do not have a complete understanding of the full data publication process, including the requirements and detailed workflows of individual stakeholders, such as the journal, author, and repository, and the touch points between them.

Bottlenecks are arising at these touchpoints, as domain and institutional repositories are becoming increasingly challenged to fit into the scholarly publication workflow. This situation is exacerbated by a current lack of awareness and communication across these relevant stakeholders. Reaching consensus on stakeholder needs and processes is a first step in addressing the challenges associated with data sharing, publication and citation, and in better aligning these workflows to facilitate Open Science.

Workshop Goals:

In this workshop we hope to discuss the detailed workflow between the journal, author, and repository that identifies the activities necessary for relevant processes to move forward in a more coordinated way. The co-chairs will share the current progress on the use case where data are submitted for preservation at or near the same time as the manuscript is submitted to a journal that will cite the data.

The co-chairs seek feedback from both journal and repository communities that will lead to recommendations for better transparency and coordination across both processes.

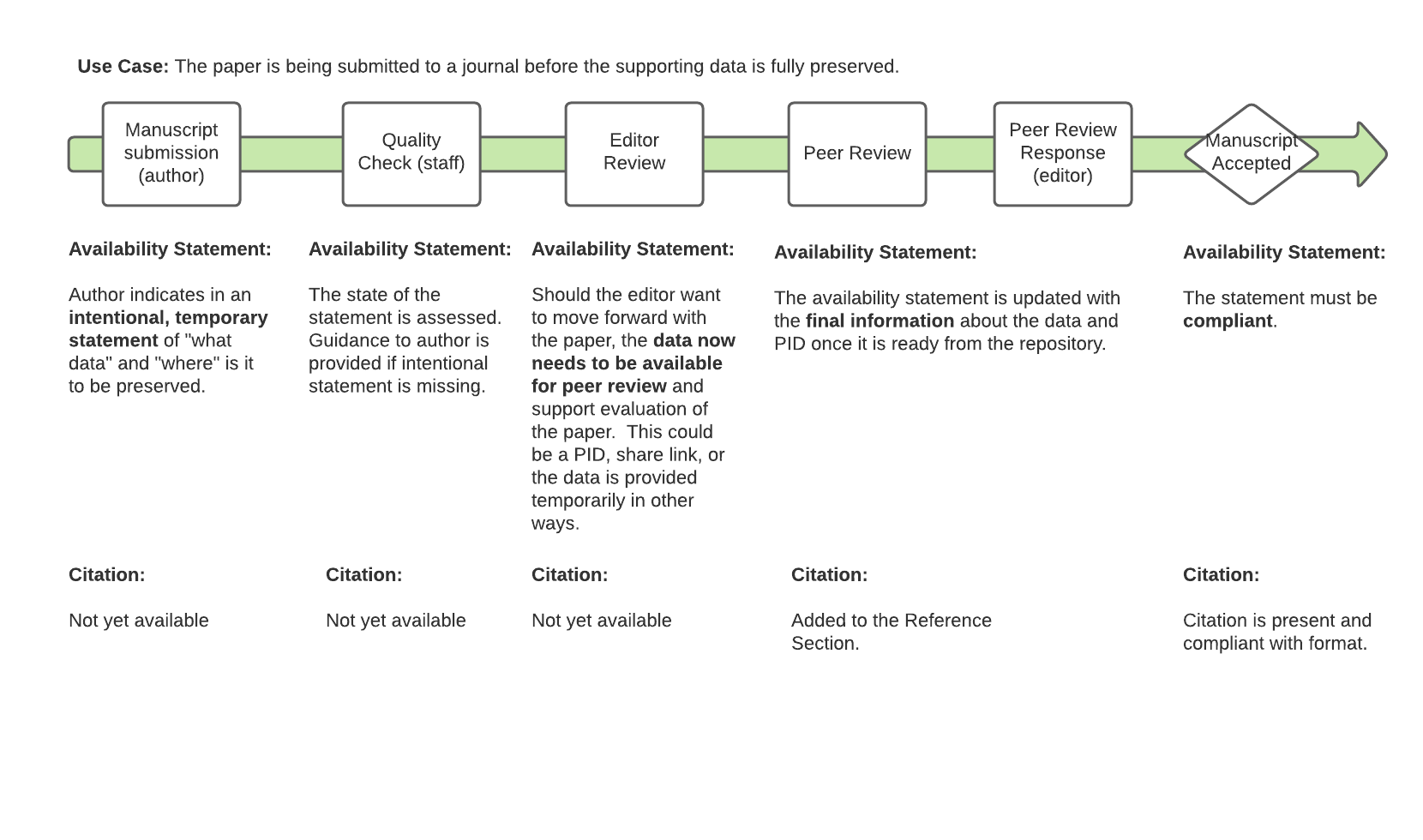
The co-chairs have presented this work for discussion at prior workshops at the Research Data Alliance and the Earth Science Information Partners.

Specific Objectives:

* Transparency of repository workflow- shining a light on the murky workflow of the journal 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.

COLLABORATIVE GROUP NOTES

Adding variations to [journal workflow activities](https://drive.google.com/file/d/1d0-xUhp_rUOJ01SaQhamepPrdZiuUoRp/view?usp=sharing):



Activities- High level View:

1. Author guidelines
2. Paper submitted (author)
3. Quality check of paper (staff)
4. Editor review
5. Peer review
   1. Daniel - the cost of HPC is not accounted for in this activity
      1. John - we don’t ask reviewers to reproduce the results
6. Final manuscript decision -- acceptance

When trying to encourage the whole corpus being published - recommend to the author that a data paper be prepared.

For repositories - a way to associate papers with datasets after both have been published. <https://support.datacite.org/docs/contributing-data-citations#>

Id: "<https://doi.org/10.7302/z23r0r29>"

{

"relationType": "IsCitedBy",

"relatedIdentifier": "10.1103/physrevlett.115.056601",

"relatedIdentifierType": "DOI"

}

],

Text used by *Nature* for Earth science papers that are being sent out to review (although this is usually edited for the specific discipline, for example to direct them to the appropriate domain repository, if that is clear):

Please note that Nature is a signatory of the FAIR data initiative in the Earth, Space and Environmental Sciences (see http://copdess.org/enabling-fair-data-project/ and https://www.nature.com/articles/d41586-019-00075-3). Consequently, if we are ultimately able to offer publication of your paper, we will require that your underlying datasets be available at the time of publication in a public repository. If this condition is unacceptable, please let me know, as we would most likely have to withdraw your paper from consideration. That said, we do recognize that exceptions exist, and that the specifics will likely need discussion. Of course, please feel free to get in touch if you have any questions.

Here is PNAS’ information for author re: Material and Data Availability (and suggestions to authors on appropriate repositories they should consider depending on data type):

<https://www.pnas.org/authors/editorial-and-journal-policies#materials-and-data-availability>

Draft recommendations for journals (to be improved):

1. Help authors select the appropriate domain/institutional/ general repository for the data (if they have not done so yet).
2. Give the author a sense of timing that the repository needs to curate and preserve the data.
3. Information on the repositories use of share links for use during paper peer review prior to paper publication.
4. 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.
5. What else?
   1. Production concerns: copyediting the data statement; ensuring appropriate reference citation (and XML tagging for both reference and the data statement!); what happens if authors change/add/delete information in the data statement (loops back to editorial office)? (Sharon, PNAS)

Draft recommendations for repositories:

1. 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.
   1. Some journals have double-blind peer review (T&F) and can only use limited number of repositories that support that.
2. 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.

|  |  |  |  |
| --- | --- | --- | --- |
| No | Name | Affiliation | Email |
| 1 | Shelley Stall | American Geophysical Union | sstall@agu.org |
| 2 | Danie Kinkade | Biological and Chemical Oceanography Data Management Office (BCO-DMO)/WHOI | dkinkade@whoi.edu |
| 3 | Lesley Wyborn | Australian National University | lesley.wyborn@anu.edu.au |
| 4 | Daniel Mietchen | University of Virginia | daniel.mietchen@virginia.edu |
| 5 | Chris Erdmann | American Geophysical Union | cerdmann@agu.org |
| 6 | Chelle Gentemann | Farallon Institute | cgentemann@faralloninstitute.org |
| 7 | Maria Shatz | NIEHS, ODS | maria.shatz@nih.gov |
| 8 | Amy Hafez | NIH Office of Science Policy | amy.hafez@nih.gov |
| 9 | Stephanie Holmgren | NIEHS, ODS | holmgren@niehs.nih.gov |
| 10 | Lisa Raymond | Woods Hole Oceanographic Institution | lraymond@whoi.edu |
| 11 | Rob Crystal-Ornelas | Berkeley Lab | rcrystalornelas@lbl.gov |
| 12 | August Muench | American Astronomical Soc. | august.muench@aas.org |
| 13 | Belinda Seto | NIH Office of Data Science Strategy | setob@mail.nih.gov |
| 14 | Sharon Chua | PNAS | schua@nas.edu |
| 15 | Elisha Wood-Charlson | Berkeley Lab (KBase, NMDC) | elishawc@lbl.gov |
| 16 | John VanDecar | Nature | j.vandecar@nature.com |
| 17 | Emma Shumeyko | PNAS | eshumeyko@nas.edu |
| 18 | Emily Boja | NCI, Office of Data Sharing | emily.boja@nih.gov |
| 19 | Vincent Fazio | CSIRO Mineral Resources | vincent.fazio@csiro.au |
| 20 | Dana Gerlach | BCO-DMO, WHOI | dgerlach@whoi.edu |
| 21 | Brooks Hanson | AGU | bhanson@agu.org |
| 22 | Susan Borda | U Michigan Library | sborda@umich.edu |
| 23 | Deb Agarwal | Lawrence Berkeley Lab | daagarwal@lbl.gov |
| 24 | Doug Schuster | NCAR/AMS | schuster@ucar.edu |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |