

Fostering the Publisher - Domain Repository Connection to Optimize the Long-term Value of Data

Kerstin Lehnert, Columbia University, Lamont-Doherty Earth Observatory
Danie Kinkade, Woods Hole Oceanographic Institution
Natalie Raia, University of Arizona
Shelley Stall, American Geophysical Union
Kirsten Elger, GFZ Potsdam
Matthew Warke, Taylor & Francis
Lesley Wyborn, Australian National University

16 Years Ago, in an Era far away?

- ▶ Data related to scholarly publications were presented within the paper as a table and/or attached as electronic supplements.
- ▶ The research community in geochemistry had just started to appreciate the value of geochemical databases for access to comprehensive, harmonized data collections.
- ▶ The compilation of these databases revealed the inadequate and inconsistent practices for reporting and documenting data.
- ▶ Database developers/operators connected with publishers to improve the quality and consistency of data and metadata reporting in the literature.
 - ▶ The 'Editors Roundtable' was born in 2007.
 - ▶ The Editors Roundtable agreed on policy recommendations in 2008 and released them in 2009.

Rationale for Building a Connection between Data Repositories & Publishers (2013)

- ▶ Scholarly publication remains a key high-value entry point for many researchers in making data available.
- ▶ The majority of data submitted along with publications are in formats and forms of storage that make discovery and reuse difficult or impossible.
- ▶ Domain repositories provide quality and standards, enriching and organizing data from multiple sources to facilitate new discoveries.
- ▶ Connection between publishers and domain repositories is essential to promulgate best practices.

COPDESS

- ▶ develop, implement, & promote guidelines and best practices for scholarly publishing with particular emphasis on the publication of data in support of open access policies.
- ▶ facilitate & foster communication and knowledge exchange among editors and publishers of Earth Science journals and Earth Science data facilities;

COPDESS: Coalition for Publishing Data in the Earth & Space Sciences; www.copdess.org

Astromaterials Data Workshop, Tucson 2021

GEOLOGY & GEOPHYSICS AGU News

✉ 🖨

🐦 f g+ ↻

Committing to Publishing Data in the Earth and Space Sciences

A new initiative joins together publishers and data facilities to enable data stewardship.

By Brooks Hanson, Kerstin Lehnert, and Joel Cutcher-Gershenfeld © 15 January 2015



Enabling FAIR Data Project



Commitment Statement of the ‘Enabling FAIR Data’ project*:




- ▶ All core research outputs (data, software, samples and sample metadata) will be directed to trusted repositories.
 - ▶ Supplements will no longer be primary “archive” for data.
 - ▶ Data are cited via persistent identifier.
- ▶ Adopt a shared set of author instructions (common set of expectations for authors in the ESES).
- ▶ Provide common expectations for publication peer review when evaluating science and determining if the data, metadata, and software are adequate.


Impact of the “Enabling FAIR Data” Project

EDITORIAL · 09 JANUARY 2019

Announcement: FAIR data in Earth science

Nature backs the Enabling FAIR Data initiative and requires authors to deposit data in community repositories.





Nature is pleased to supi

SPRINGER NATURE

[PDF version](#)

RELATED ARTICLES

Google unveils search engine for open data 

China backs bold plan to tear down journal paywalls 



“AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

- Depositing the data and software in a trusted repository, as appropriate, and preferably with a DOI*
- Including an [Availability Statement](#) as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software*
- And including [citation\(s\)](#) to the deposited data and software, in the Reference Section.*

Astrom

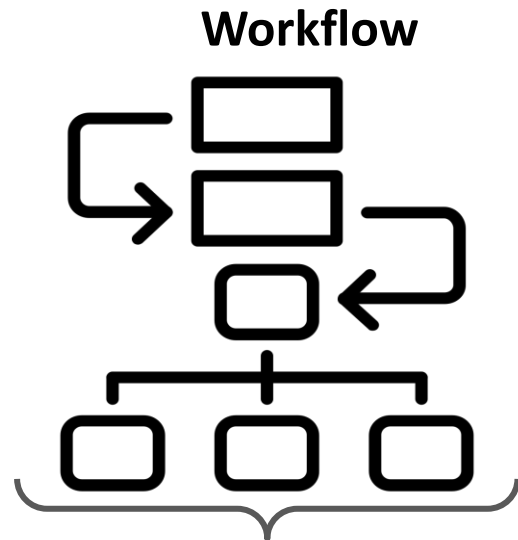
“...As such, from January 2019, we will require that authors of papers in Earth sciences (and related space and environmental science fields) make supporting data available to others through [community repositories](#) where available.”

2021: Not There Yet

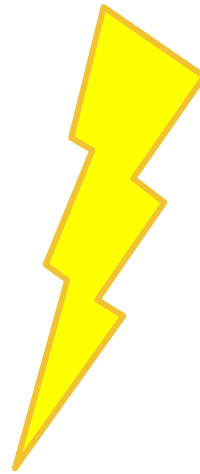
- ▶ Implementation of FAIR Data principles by publishers has revealed the lack of clear and detailed guidelines, workflows, and policies that help authors, editors, and data curators to navigate the new publishing requirements.
- ▶ Majority of data in geo/cosmochemistry still in tables or supplements of articles - many journals have not or not fully implemented FAIR data.
- ▶ Data still lack internationally endorsed and governed best practices and technical standards.

The Challenge: Coordination of Data Preservation with Paper Peer-Review Process

Domain Repository Data Publication

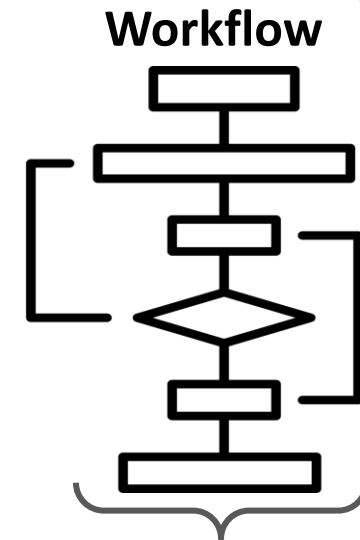


Author's Published Datasets



Friction!

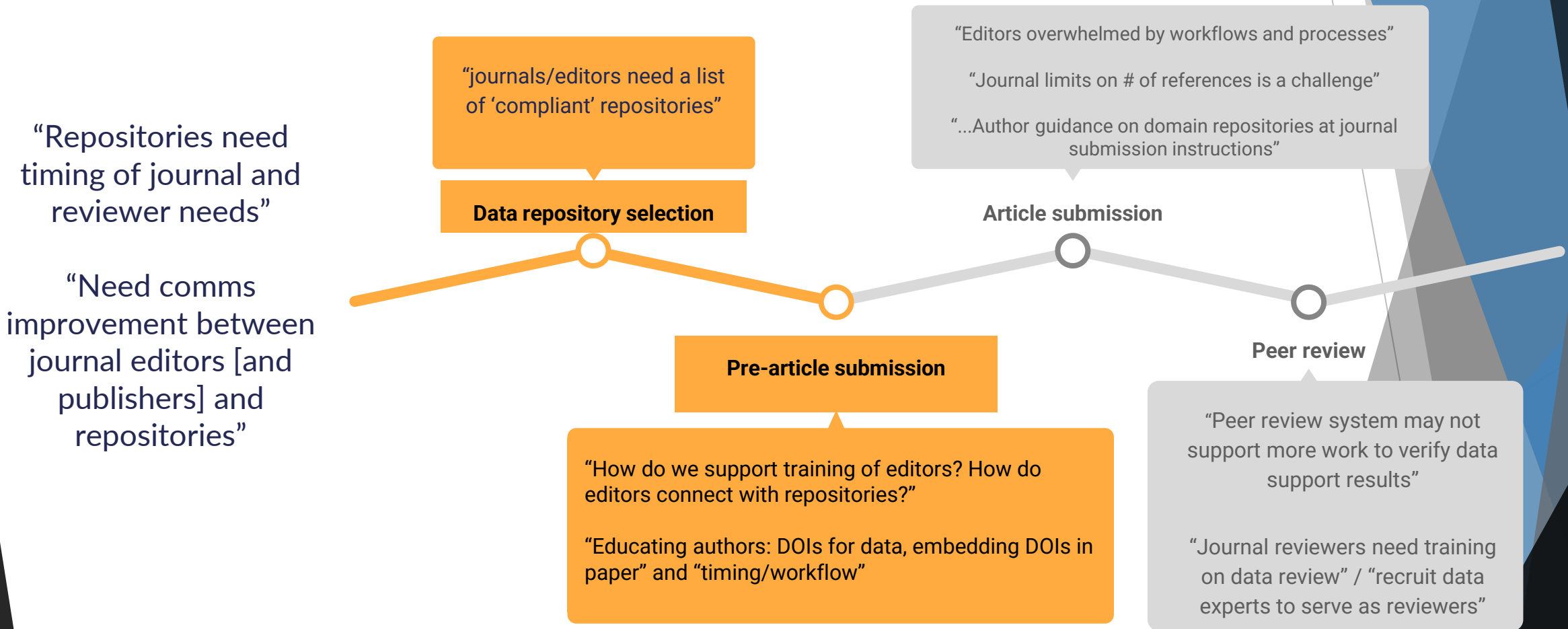
Journal Manuscript Publication



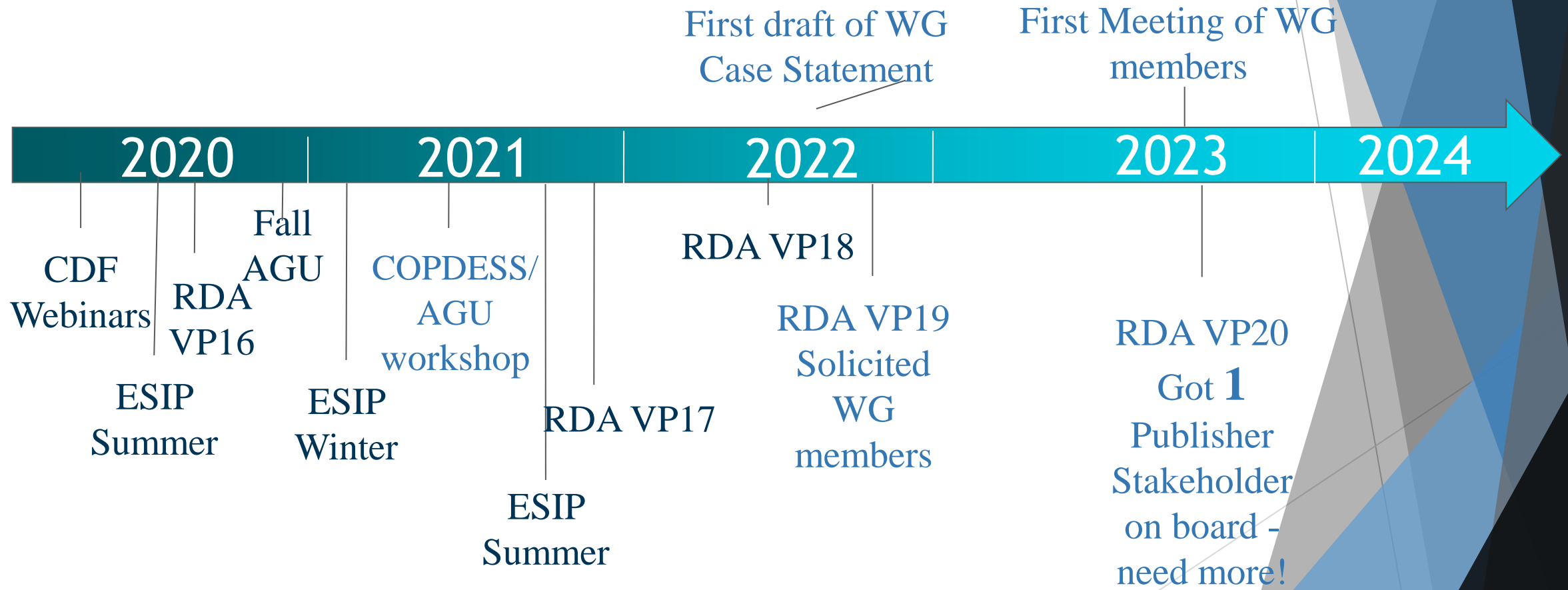
Author's Scholarly Publication
with Dataset DOIs

A new RDA WG working group has been established to reduce friction currently present between repository and publisher workflows as they pertain to- and rely upon the data publication process.

The Challenge: Coordination of Data Preservation with Paper Peer-Review Process



History: Coordination of Data Preservation with Paper Peer-Review Process (“Repo2Pub”)



2024 Status: RDA Working Group Established!

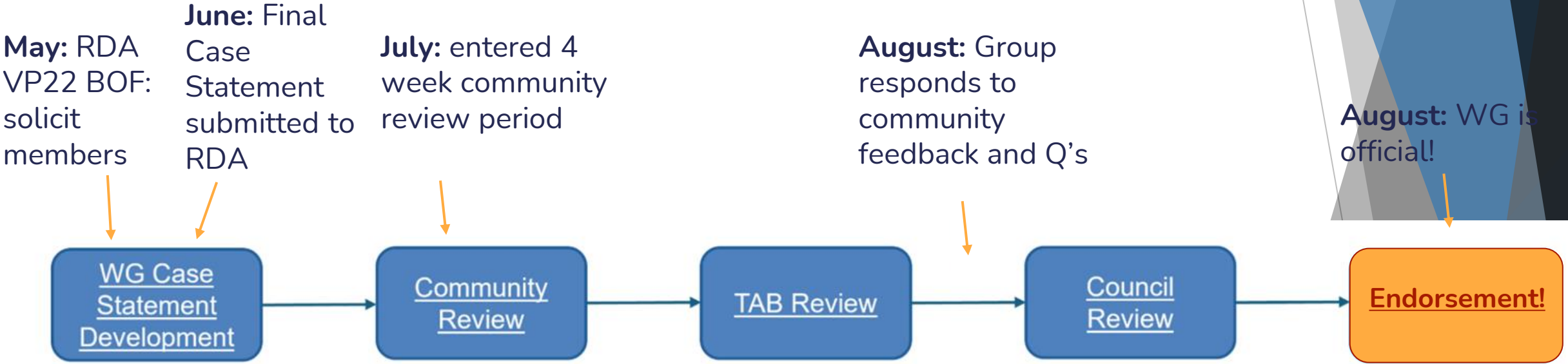


Figure 1: WG Case Statement Review Process – from Case Statement Development to Endorsement.¹

Repo2Pub WG: Value Proposition

- ← The Repo2Pub WG aims to ease repository bottlenecks, facilitate data review, and increase transparency for the authors.
- ← Bring together key stakeholders
- ← Build on and extend earlier RDA & COPDESS work
- ← Establish a common understanding of stakeholder workflows and needs
- ← Reach consensus to develop recommendations necessary to streamline data publication.

Repo2Pub WG: Target Adopters

- ▶ **Repository Managers:** Through repository policy, guidelines and education (as facilitated through organizations such as US CDF, ESIP, RDA, WDS and EGU/AGU engagement)
- ▶ **Journal Publishers/Editors:** Through journal policy, author guidelines, author education (facilitated by societies such as AGU/EGU and COPDESS)
- ▶ **Secondarily:** researchers, funders

Repo2Pub WG: Objectives

- **Establish a shared understanding** of ESES repository requirements **AND** journal needs concerning data availability
- **Document a set of use cases demonstrating the current (somewhat disconnected) AND preferred steps** in the publication processes
- **Educate and inform stakeholders on the touchpoints with the author** in respect to ESES data that underpins a scholarly publication.
- ✓ **Create a visual model of publication workflow** that includes journals and discipline-specific repositories (i.e., refine the current draft individual workflows)
- ✓ **Create a set of recommendations for both** journals and repositories (i.e. refine current draft set of recommendations).

Repo2Pub Working Group: Timeline

Use Case
Template
Development



- ▶ Months 1-6: Document a set of repository use cases that demonstrate bottlenecks in publication process
- ▶ Months 6-12: Creation of a generalized visual model of publication workflow for journals and detailed data workflow for repositories based upon repository use cases.
- ▶ Months 12-18: Creation of a set of recommendations:
 - ▶ For journals, containing actions that support and facilitate repository workflows.
 - ▶ For repositories, containing actions to better support journal manuscript publication process needs.

Repo2Pub & CDF: Ways to Get Involved

- ▶ Join the RDA Repo2Pub Working Group page to receive updates on upcoming meetings and progress: <https://www.rd-alliance.org/groups/coordinating-earth-space-and-environmental-science-data-preservation-and-scholarly-publication-processes/>
- ▶ Share your challenges & needs via the current use case collection effort!: Natalie Raia, nraia@arizona.edu (to volunteer for use case contribution or answer questions via email, set up short Zoom call with you + your colleagues, etc.)