National Academies of Sciences, Engineering, and Medicine's

Board on Research Data and Information

Toward an Open Science Enterprise

September 18, 2017

Shelley Stall, Director Data Programs

Brooks Hanson, Sr. VP Publications

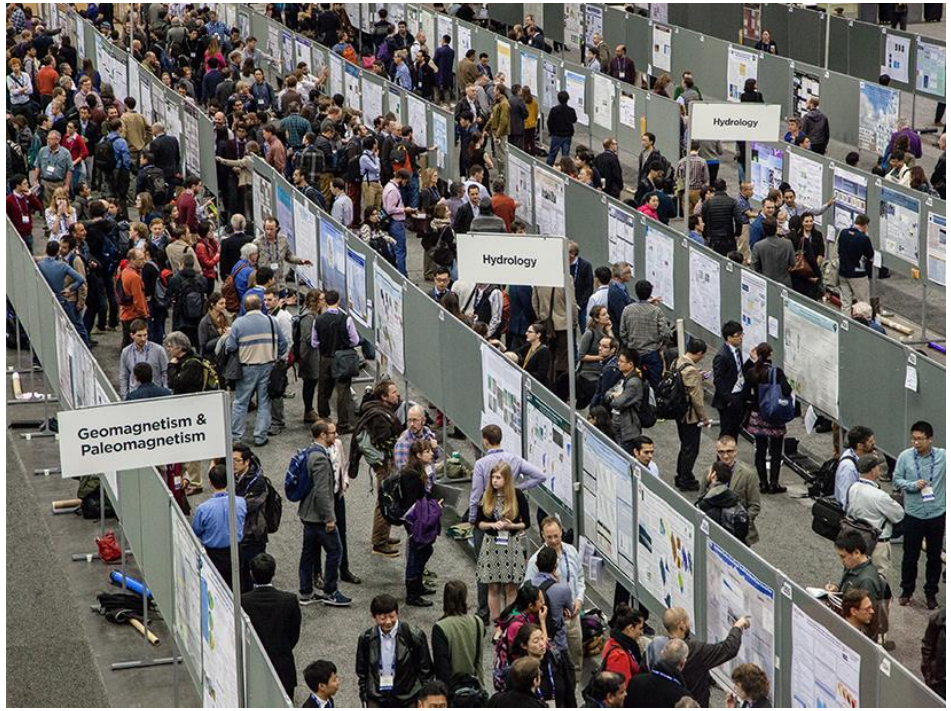
American Geophysical Union

About AGU

- Largest Earth and space science (ESS) society
  - 60,000 members; 137 countries
  - Much more than “geophysics”
  - Large annual meeting, 25,000 attendees

- Largest Society Publisher in the ESS
  - 20 peer-reviewed journals
  - >6000 published papers in 2016

- Outreach to government leaders and the public



AGU Publication Open Practices

- 20 Peer-reviewed Scholarly Journals
  - 4 Gold Open Access – since 2013
  - 16 Open Choice/Hybrid

And Expanding…

- All Commentaries — Open
- All Supplements — Open
- All References — Open (through Crossref)
- All Publications highlighted in social media — Open
- All Publications, after 24 months — Open
  - Authors may self-archive (Green Open Access) after 6 months
- All EOS.org articles — Open

- CHORUS, Crossref, DataCite Participation

AGU's position statement on data affirms that

"Earth and space sciences data are a world heritage. Properly documented, credited, and preserved, they will help future scientists understand the Earth, planetary, and heliophysics systems."

https://sciencepolicy.agu.org/files/2013/07/AGU-Data-Position-Statement-Final-2015.pdf

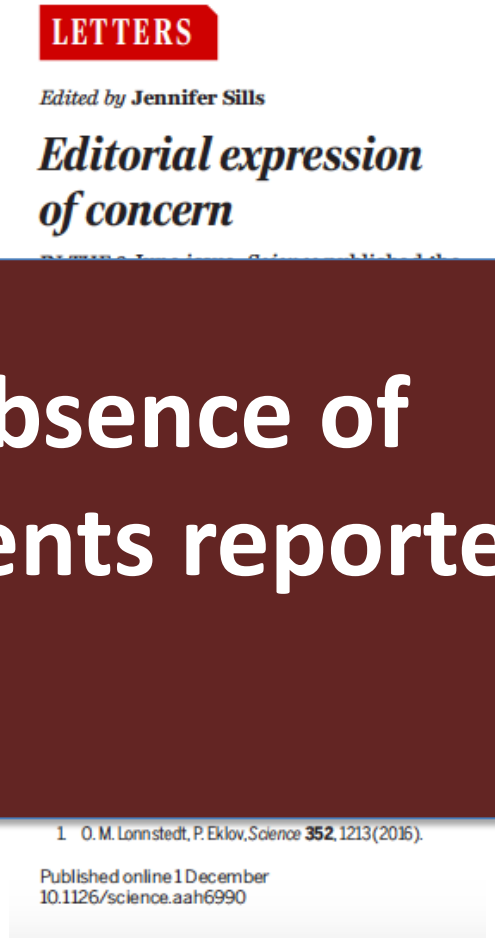
Photo Credit: Photo by Rick Meyers on Unsplash

AGU's Publication Data Policy states...

All data necessary to understand, evaluate, replicate, and build upon the reported research must be made available and accessible whenever possible.

Open For Interpretation

TRUE STORY – Dec 1, 2016



LETTERS

Edited by **Jennifer Sills**

Editorial expression of concern

In the 3 June issue, *Science* published the Report "Environmentally relevant concentrations of microplastic particles influence larval fish

Retracted May 3, 2017 - absence of original data for the experiments reported in the paper;

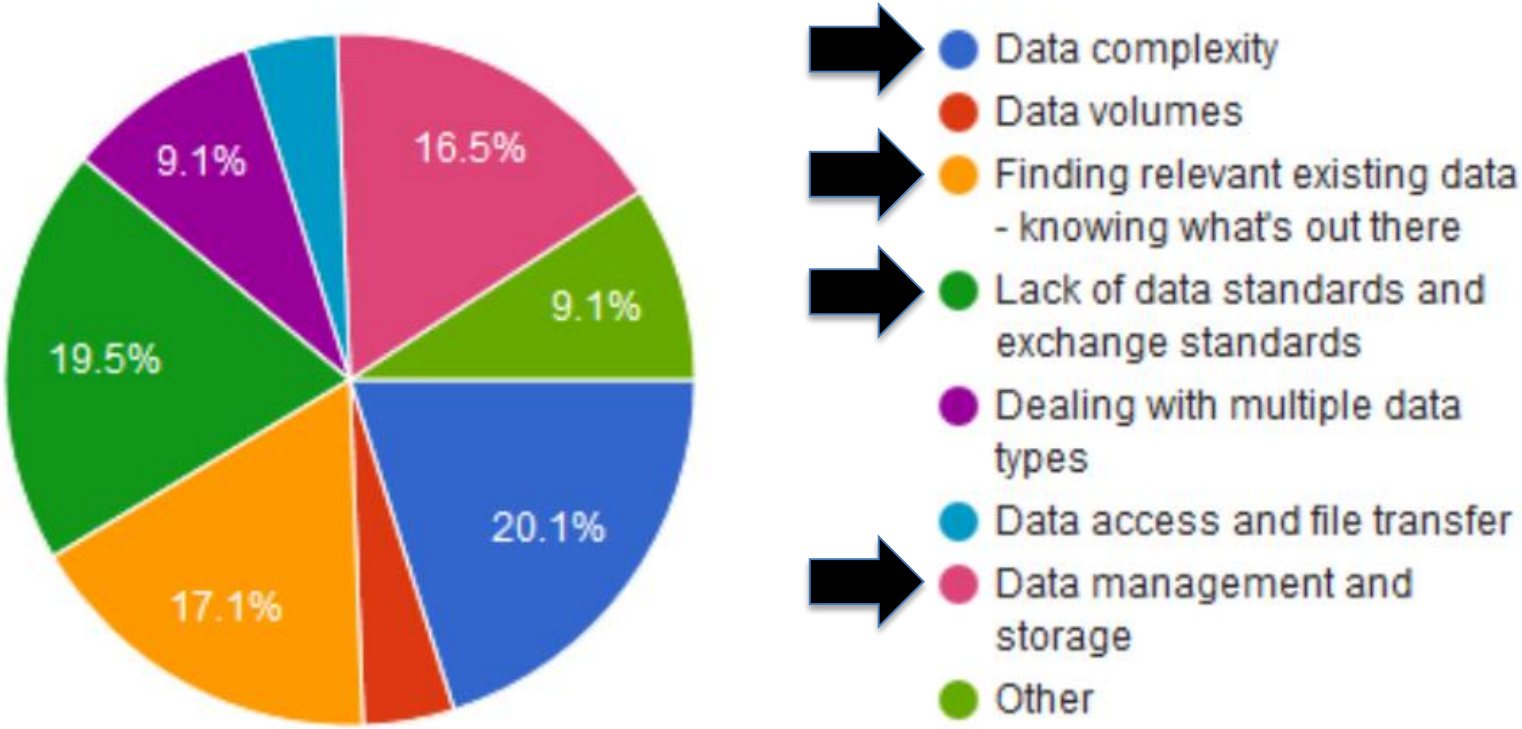
to understand, assess, reproduce, or extend the conclusions of the paper.

1 O. M. Lönnstedt, P. Eklov, *Science* **352**, 1213 (2016).

Published online 1 December
10.1126/science.aah6990

Researcher Challenges with Data Use

The top four issues accounted for 73% of respondents



As a Researcher, you need…

Findable

- **Publications to include the data citation** that identifies the repository where the data is located. Use of persistent identifiers.
- Same for any relevant software.

Accessible

- The repository to be **freely accessible**.

Interoperable

- **The data to be in a format appropriate** for the data domain. If multiple formats are used, the repository should have the ability to provide the data in all the accepted formats through conversion.

Reusable

- **The data in the repository to be well documented** – allows you to determine if it's "fit for use" without having to contact the PI.
- Clear information on the **data usage license**.

FAIR Guiding Principles

Findable
Accessible
Interoperable
Reusable

Developed by Force11.org

Article in Nature: Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018 doi: 10.1038/sdata.2016.18 (2016).

As a Repository supporting the Research Data Lifecycle…

Findable

- **Web accessible data services** that allow discovery tools to locate your data holdings.
- Persistent Identifiers (e.g. **Digital Object Identifiers** – DOI) for all data sets.
- Landing pages that support **data citations.**

Accessible

- **Tools** for researchers to access and investigate data sets that are of interest.
- Free access to holdings.

Interoperable

- **Metadata standards** and standard data set formats.
- Standard vocabularies

Reusable

- Clear and accessible **data usage license.**

As a Journal, Custodian to the Scientific Record...

All data necessary to understand, evaluate, replicate, and build upon the reported research must be made available and accessible ~~whenever possible.~~

Publishers and Repositories are Working Together…

- TOP (Transparency and Openness Promotion) guidelines, signed by 2900 journals and organizations
- COPDESS.org (Coalition on Publishing Data in the Earth and Space Sciences)—Statement of Commitment endorsed by most publishers and repositories in the Earth and space sciences
- Joint Declaration of Data Citation Principles endorsed by 114 organizations including most major publishers.
- Reproducibility conferences and outcomes (AAAS and other orgs)
- Quality/certification standards for repositories expanding

Challenge is practicing what you preach

Coalition on Publishing Data in the Earth and Space Sciences (COPDESS.org)

Connecting Earth Science publishers and Data Facilities to help translate the aspirations of open, available, and useful data from policy into practice.

- Formed in October 2014
- Endorsed a **Statement of Commitment, 2015**
- Includes: joint best practices between journals and repositories; references.

New Grant from Laura and John Arnold Foundations (LJAF)

Develop best practices and standards

that **will connect** researchers, publishers, and data repositories in the Earth and space sciences

to enable **FAIR** data.

This will accelerate scientific discovery and enhance the integrity, transparency, and reproducibility of this data.

This project will help:

1) researchers understand and follow expectations regarding data management and metadata required for publication

2) publishers adopt and implement common standards and best practices around datasets, metadata, acceptable repositories, and citation supporting publication

3) repository recognition of their valuable role in data lifecycle providing curation services, persistent identifiers, landing pages, data discoverability, free access, and clear data usage licenses (maximum potential for reuse, e.g. CC0)

Enabling FAIR Data – Key Points:

- Community-driven solution with AGU as convener
- Builds on the work previously done by COPDESS.org
- Data associated with publication will be open "by default"
- Quality of data documentation (metadata) becomes consistent – supports FAIR principles
- ESS Publishers and Repositories adopt project recommendations and guidelines

Open Data Issues -

- Not all Data Types have a Domain Repository with the mission to curate and preserve
- Not Enough Repositories for Trans-Disciplinary Data
- Need to Scale Up existing Data Management Training for Researchers on Best Practices and Tools to be Available to All Researchers
- Need Recognition of Data Products as a First Class Research Product
- Need Recognition of the Data Manager/Curator as a Valuable Contributor to the Scientific Record

Open Data - Take Aways

- Data infrastructure is developing.
- Aspiration of Open and FAIR Data is achievable in the near future.
- Long term funding model for repositories needs to be developed to adequately supporting open data and preservation.
- Incentives for the researcher to better manage, document, and share their data need to be developed.

Questions?

Contact Information:

Shelley Stall
Director, AGU Data Programs

Enabling FAIR Data Project Status:

www.copdess.org select “Enabling FAIR Data Project” from left menu

@ShelleyStall
sstall@agu.org