CODATA:
Data Citation Workshop Perspectives from Editors and Publishers

Brooks Hanson
Director, Publications, AGU
bhanson@agu.org

Requiring access to data is not new…

- AGU position statement on data in 1997; first publications data policy in 1993.

- *Science*, *Nature*, *PNAS*, and other journals required protein structure deposition in late 1990's; use of Genbank and other established domain repositories soon afterward.

  - *Science* formalized and strengthened data policies in ca. 2000 and 2005 and began requiring access to code and strengthened further in 2011 (see 11 Feb. 2011 special issue of *Science* on data).

- OSTP mandate in 2013; similar requirements emerging globally.

What is New, and What Publishers and Partners Are Doing Now...

- Most major publishers and funders are <u>really</u> requiring access to data.
- More are requiring access to code, but not <u>really</u> as much
  - Enforcement is starting to occurring
  - Devil is a bit in the details on both
- Implementing common best practices around citing data and recently software
- Developing other common best practices around data and identifiers
- Various efforts emerging to extend awareness to researchers of these best practices and help implement them in workflows.
- Interactive online manuscripts and editorial workflows that include data dynamically and in development.
- Efforts to help repositories improve their best practices

@AGU

What are best practices...

- Joint declaration of data citation principles
  - Citing data in the references as a reference
  - Separate data publications when appropriate (more journals now available)
- Transparency about how researchers can access data (e.g., statement in acknowledgements)
- Include ORCID's and other community identifiers
  - Funders, samples, author-credit, institutions (still to come)
- Use trusted domain repositories if they are available
- Use repositories that allow for data access during peer review
- Supplements should follow NISO guidelines
- All references should be in main reference list (not in supplements)
- Key references and data should be available at time of publication (no unpublished or in-press references)

AGU

Publishers and Repositories are Working Together...

- **TOP** (transparency and openness promotion guidelines) 538 journals

- **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 109 organizations including most major publishers.

- Reproducibility conferences and outcomes (AAAS and other orgs)

- Certification standards for repositories

Challenge is practicing what you preach

AGU

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.

- *See Kerstin Lehnert's talk this afternoon*

Next Steps

- Software citation standards released by FORCE11 and aligning with several other efforts; *look for report from recent AAAS meeting on software reproducibility*.

- Some major publishers requiring ORCIDs in 2016 (PLOS, eLife, Royal Society, IEEE, AGU, AAAS...)

- AGU started two programs: Data Management Maturity (for helping repositories adopt best practices) and Training in Data Science (for researchers)--More info from **Shelley Stall** and Cyndy Chandler. "Data Fairs" at our Fall Meeting.

- Starting a group associated with RDA to help journals implement best practices. Others also engaging in this outreach.

- Publishers exploring working together on messaging to authors and in author workshops around data.

The details about data…

- What data are required (do you really mean all of it)?
  - Usually determined by the discipline and the data that are typically stored in domain repositories. More rarely the raw data, but certainly the tabulated data in support of reported results. Anonymized where appropriate.
- Enforcement of policies
  - We need to help authors (starts at funding agencies and data collection), and be aware (acknowledgement statement). It is an ethical obligation and key for advancing science
  - Ask reviewers and editors about data availability.
  - May require editorial statements of concern after publication if data are not provided.
  - Hold or coordinate publications until data and references are available.
- How about someone stealing my data or scooping me?
  - Widely adopted citation standards and norms
  - Some communities have much experience in open data
  - Some urban legends used as excuses
- I received data from someone else or it is commercial or restricted by my government or laws and I can't release it.
  - Transparency in access.
  - IP is ok if data, data products, or software are available and scholarly reuse is allowed.
  - Alert authors to think about data and access up front, in negotiating transfer agreements.

8

@AGU