Implementing Best Practices around:

FAIR Data

In Scholarly Publishing

November 2017

Brooks Hanson and Shelley Stall
bhanson@agu.org

Enabling FAIR data in the scholarly literature

- Embrace that published papers are only part of the research
- They must contain useful and reliable 2-way links and identifiers to other secure resources for integrity and discoverability:
  - Context (metadata) around these links are critical
  - Data, software, repositories, samples (IGSN)
  - Funding information
  - Author information (ORCID, CREDIT, institutions)
  - Reference information (semantic context is coming)
- We need efficient ways to help authors, publishers and repositories preserve these links: Standard, expected, sensible, easy.

Recent Alignment by Publishers, Repositories, and Funders Around Best Practices

- **TOP** (transparency and openness promotion guidelines)--2900 journals
- **COPDESS.org** (Coalition on Publishing Data in the Earth and Space Sciences)—Statement of Commitment--most publishers and repositories in the Earth and space sciences
- **Joint Declaration of Data Citation Principles**–114 organizations
- **Software Citation Principles**: https://doi.org/10.7717/peerj-cs.86
- **Reproducibility conferences** and outcomes (AAAS and other orgs)
  - Best practices around: Clinical trials, Lab studies, Field data, Software, Industry-academic research
- **Authorship** (https://doi.org/10.1101/140228 submitted to PNAS)
- Quality/certification standards for repositories

Challenge is practicing what you preach

Current status:

- Publishers increasingly requiring data (and code) availability
  - Supplements still being heavily used (no metadata, pdf often)
  - Growing use of repositories, domain and general (Figshare, Dryad, institutions).
  - Few standards on metadata or linking (limiting discoverability; interoperability).
  - “Available from authors (yeah, right)” still common
  - “unpublished” references still commonly allowed
- Best practices for FAIR data are available
- Great examples in some disciplines/repositories of successful implementations and solutions but not widely adopted.
- Leverage and scale these solutions!!!

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

Software: Current Status

- Leading journals have software transparency standards
- Community best practices emerging
- But…little uniformity in those best practices and limited awareness among authors, editors

- Key issues:
  - Licenses—Use MIT or other software license, not CC-BY (which require attribution and documentation of any changes)
  - Github has limited metadata (can use zenodo as a landing page).
  - IP
  - Developing common standards and researcher expectations

Recommendations

- Share data, software, workflows, and details of the computational environment that generate published findings in open trusted repositories.
- Persistent links should appear in the published article…
- To enable credit for shared digital scholarly objects, citation should be standard practice.
- To facilitate reuse, adequately document digital scholarly artifacts
- Use Open Licensing when publishing digital scholarly objects.
- Journals should conduct a reproducibility check as part of the publication process and should enact the TOP standards at level 2 or 3
- To better enable reproducibility across the scientific enterprise, funding agencies should instigate new research programs and pilot studies.

PeerJ CompSci

✔ PEER-REVIEWED

Software citation principles

Digital Libraries | Software Engineering

Arfon M. Smith *1, Daniel S. Katz ✉ *2, Kyle E. Niemeyer *3,
FORCE11 Software Citation Working Group

September 19, 2016

https://doi.org/10.7717/peerj-cs.86

- **Importance:** Software citations should be accorded the same importance as … publications and data
- **Credit and attribution:** Software citations should facilitate giving scholarly credit and normative, legal attribution to all contributors, recognizing that a single style or mechanism of attribution may not be applicable to all software.
- **Unique identification**
- **Persistence**
- **Accessibility:** should facilitate access to the software itself and to its associated metadata, documentation, data, and other materials necessary for both humans and machines to use of the software.
- **Specificity:** should facilitate identification access to, the specific version of software that was used..

Authorship best practices: McNutt *et al.* https://doi.org/10.1101/140228 (BioRxiv and submitted to PNAS)

- ORCID
- CREDIT
- Revised authorship criteria
- Responsibilities for corresponding authors
  - Increased transparency

Office of Budget Finance & Award Management (BFA)
Office of Budget, Finance, & Award Management
Budget Division
Division of Acquisition and Cooperative Support
Division of Financial Management
Division of Grants & Agreements
Division of Institution & Award Support
Large Faci
Advisory C
Business a
Committee
External L
Chief Financial Officer Council

Data Management & Sharing Frequently Asked Questions (FAQs) -

UPDATED NOVEMBER 30, 2010

1. What constitutes "data" covered by a Data Management Plan?
2. Is a plan for Data Management required if my project is not expected to generate data or samples?
3. Am I required to deposit my data in a public database?
4. There is no public database for my type of data. What can I do to provide data access?

Not necessarily. The expectation is that all data will be made available after a reasonable length of time. However, what constitutes a reasonable length of time will be determined by the community of interest through the process of peer review and program management.

9. Does this policy mean that I must make my data available immediately, even before publication?

NIH Grants Policy Statement
Notices of Policy Changes
Compliance & Oversight
Select Policy Topics +

Frequently Asked Questions

Data Sharing

Public FAQs
NIH Staff FAQs

Last Revised: February 16, 2004

Filter Refresh

A. March 5, 2003

Expand/Contract

1. Why should I share my final research data?
2. Who benefits from data sharing?
3. Is data sharing widely accepted as a good practice?
4. What do you mean by final research data?
5. Does "final research data" include data that were not originally produced under an NIH grant or contract?
6. What do you mean by unique data?
7. What kinds of data are candidates for sharing?

Grant from Laura and John Arnold Foundations (LJAF)

- **Align/develop best practices and standards across the Earth and space sciences to enable FAIR data**

- Develop common solution(s) for researchers, publishers, editorial systems, and data repositories

Community-Driven Project – Partnership Includes:

- **Science Data Communities**
  - AGU
  - Earth Science Information Partners
  - Research Data Alliance (RDA)
  - COPDESS
  - Earthcube/CDF
  - DataCite
- **Publishers**
  - AGU
  - *PNAS*
  - *Nature*
  - *Science*

- **Repositories and COPDESS Signatories**
  - National Computational Infrastructure (NCI)
  - AuScope
  - Australian National Data Service
- **Infrastructure**
  - Center for Open Science

And Growing!!

The Goal:

Publishers will adopt common standards in their editorial systems around workflows, datasets, metadata, acceptable repositories, and data citation. Will connect w/ repositories via api's for efficient metadata exchange.

Repositories will adopt standards and best practices around, persistent identifiers, landing pages (for exposing metadata), access, embargoes (for peer-review), data citations, and licenses.

Researchers will know what to expect across all journals and be able to prepare data early in workflow.

Timeline – 18 Months

Preparation for First Stakeholder Meeting	Aug 1, 2017 – Nov 15, 2017
First Stakeholder Meeting (tomorrow)	Nov 16 – 17, 2017
- Working Groups Formed and Active	Nov 17, 2017 – Apr 2018
- Development of Guidelines, Recommendations, and Policies for Journals and Repositories	Nov 17, 2017 – Apr 2018
- Testing of Guidelines, Recommendations, and Policies	Apr 2018 – June 2018
Second Stakeholder Meeting	June 2018
- Adoption and Implementation of Guidelines, Recommendations, and Policies	June 2018 – Feb 2019

How To Participate…

- Stay Informed and help inform your researchers and colleagues:
  - http://www.copdess.org -> Enabling FAIR Data Project
- Participate in the Stakeholder Alignment Survey – November 2017
- Participate in a Working Group
  - Formation is during First Stakeholder Meeting – Nov 2017
- Support FAIR Principals In the Rest of the Lifecycle
  - Incentives
  - Communication
  - Alignment

Larger Effort Needed

- Support and publicize these community efforts around best practices
- Publishers need to follow current best practices
  - Get references out of supplements in online versions (all references in main text); open up references at Crossref
  - Help authors (include data best practices and expectations into workshops, instructions…)
  - Ensure integrity (no unpublished references; data availability statements).
- Societies should recognize data stewardship in awards and recognition (fellowship), specifically.
- Funders need to standardize DMP's and follow through on these
  - Update guidelines and FAQs to follow best practices
  - Support leading publishers
  - Support leading repositories
- Implement identifiers fully (affiliations and repositories)