Open Data for Library Publishing

Monica Westin
Publications & Engagement Manager,
California Digital Library
Outline of presentation

1. Data across the disciplines
2. Data publishing best practices
3. Why open data sharing?
4. Supporting data sharing
5. Educating journal editors & authors
6. Data publishing in the CDL publishing program
Definition of research data

The necessary information to evaluate, re-use, or reproduce your research.
Data across disciplines

Ex: Art history

- Images
- Spectral Imaging
- High Resolution Scans of Brush Strokes
- Documentation
- Measurements & historical tracking related to art & artists

Art history data set courtesy of K. Bender
Data across the disciplines, continued

- Linguistics: Video, Audio, Transcripts
- Stats: Tabulated Data, Software, Algorithms
- City Planning: Geographic base map files
- Biological Sciences: values extracted from images, blots, gels
- Quantitative Social Sciences: Transcripts, tabular data, imaging
Supplemental files in articles for data sharing are inadequate!

Why? No DOI, little to no metadata, difficult to discover independently of the article, often converted to PDFs...
Data Publishing: the gold standard of data sharing

= the process of making research data publicly available, reusable, and reproducible
3 core features of a true data publication

1. Data are publicly **accessible** and **preserved** indefinitely
2. Data are described: data have **metadata**
3. Data are **citable** and **credible**
1. Data are publicly available & preserved indefinitely

A **stable repository** can ensure data will be available and stored for an indefinite amount of time.

**Sensitive data** should be stored in a **preserved and compliant space**.

[re3data.org](http://re3data.org)
2. Data are described: data have metadata

Metadata for data range from information regarding software used for analysis to who funded the work.

The Digital Curation Centre provides a disciplinary metadata guide that lists metadata standards by discipline. Some general dataset standards that are discipline agnostic include DataCite, Project Open Data, and DDI.
3. Data are citable and credible

Data need to have a **persistent identifier** (a stable link) that can be referenced and cited.

** Citation **

Fitzgerald, Clark; Zhang, Michael (2018), Caltrans PEMS highway sensor average flows by occupancy, v3, UC Davis Dash, Dataset, [https://doi.org/10.25338/B8QC7F](https://doi.org/10.25338/B8QC7F)
Why is data sharing important?*

* to your journal editors and the authors they work with
Carrots: for the good of research & advancement

In the Age of Trump, Open Science Is Crucial

With funding cuts to major science agencies looming, it is now more important than ever for researchers to embrace transparency and data sharing.

03.30.2017 / BY Marcus Banks

73% of academics say access to research data helps them in their work

Most scientists 'can't replicate studies by their peers'

Table 1: Reproducibility of research findings
Preclinical research generates many secondary publications, even when results cannot be reproduced.

From
Drug development: Raise standards for preclinical cancer research
C. Glenn Begley & Lee M. Ellis
Nature 483, 531–533 (29 March 2012) | doi:10.1038/483531a
Sticks: mandates and policies

NIH Data Sharing Policies
This table lists data sharing policies in effect at NIH.

BILL & MELINDA GATES FOUNDATION OPEN ACCESS POLICY

5. Data Underlying Published Research Results Will Be Accessible and Open Immediately.

PLOS ONE To Require Public Access for Data
Support data publishing in your program

1. **Data repository** for your journal’s authors to submit to if there isn’t an appropriate disciplinary repository (e.g. campus data repository)
2. **data availability statement** or other way for authors to share the location of data in the article metadata
3. appropriate display of the **data related metadata**
4. **educated journal editors** who will implement data publishing best practices...
Journal editors need to work with...

Their **editorial boards** to establishing guidelines, procedures, and even full policies

**Reviewers** and **managing editors** to understand how to act on these policies

**Authors** to set appropriate expectations starting at the point of submission
Supporting open data policy authorship

- Author guidelines with sample text & instructions
- Instructions for reviewers
- Guidance on sensitive, proprietary, and private data
- Information on data licensing
- Sample data availability statements

*Consider liaising with data curation librarians, research data specialist, or other data experts in the library for data publishing policy support.*

*California Digital Library research data specialists*
What we’re doing at CDL: Repository data availability statement

Data Availability
- Public repository

https://dash.ucop.edu/stash/dataset/doi:10.6071
What we’re doing at the CDL: workflows

**Current**: Encouraging best practices on the systems that we currently have (working around OJS supplemental file system)

**Future**: Fully realized feature to integrate eScholarship publishing workflows with a Data Availability Statement section in journal metadata that will integrate flexibly with data repositories
Open questions for the library publishing community

- What are some of the **challenges** that we face as we work to implement best practices for data publishing into our library publishing programs?
- Is it time for library-based publishing platforms to create **basic criteria around data publishing**?
Thank you!
Monica Westin
Publications & Engagement Manager, California Digital Library
monica.westin@ucop.edu

References:

