Data Publishing

Where researchers are and how libraries are moving the needle

Ariel Deardorff, MLIS
Data Services Librarian
University of California, San Francisco Library

About me

Data Services Librarian

University of California, San Francisco

Perspective from a biomedical research institution

Open science researcher

The ideal scenario

"Hi, I am a faculty member starting a new research project. What should I do now to plan for data publishing?"

What actually happens

"Hi, I am submitting an article to PLOS One and it says something about data sharing. What does that mean and how do I do it ASAP?"

What's going on here?

Many researchers don't know about data publishing, or they don't know how to comply

What are the barriers to data publishing?

First, the term "data publishing" is confusing

———

- Many researchers don't really understand what "data publishing" means
  - Think that publishing = articles
- We sometimes say "data sharing" but that has other connotations:
  - Researchers think of sharing with data with their colleagues or collaborators
  - Universities often think of data sharing in terms of sharing data with industry or other universities

Researchers don't know what data to publish

———

Q: What data should I be sharing?

- They are confused about what version of their data to share
  - Should they be sharing images, cleaned data, raw data, graphs?
- Often forget about necessary metadata/documentation that needs to accompany data
  - Ex: data dictionaries

Many researchers didn't get participant consent to share

———

Q: But I said in my consent forms I would destroy the data!

- For those working with human subjects, many researchers have used older consent forms that mention destroying data
- Going forward they can update consent forms but sometimes need to re-consent current participants

Others aren't sure about de-identification

———

Q: Will the data repository de-identify my data for me?

- De-identification can be challenging, researchers are worried they will do it wrong
- Especially tough in medical settings where patient privacy is very important and fines for a data breach can be huge.
- Uncertainty around qualitative data

Most don't know where to publish it

———

Q: Can people just email me and ask for my data?

- Many researchers are not familiar with data repositories
- Need help locating appropriate repositories in tools like FAIRsharing or re3data

Some don't have appropriate repositories for their data

___

Q: Where should I put my clinical data?

- Still lacking infrastructure for some types of data
- Not a go-tool place for clinical data
- Not as many options for managed access sensitive data repositories

Others don't see the value

———

Q: But it's my data!

- Some researcher worry about being scooped
- Many feel ownership over their work and don't want to "give it away for free"
- Need a gentle reminder that their research is publicly funded

Most researchers don't want to spend the time

———

Q: How long does this take?

Fact: Preparing useful data takes time!

How are libraries moving the needle?

We train current and future researchers

———

- We teach data management classes at the Library (informed by FAIR data principles)

    - Ex: [UCSF Secure Research Data Management Class](#)

- We partner with faculty to teach classes in the undergraduate and graduate curriculum

    - Ex: UCSF Data Management class in Responsible Conduct of Research for Graduate Students and Postdocs

- We teach classes and give updates as part of departmental meetings and grand rounds

We give advice about data publishing

———

- Librarians meet 1:1 with researchers
- We come to labs or teams to provide group support
- We are great at connecting researchers to other resources on campus
  - Ex: At UCSF we are constantly connecting folks with our de-identification service

We curate data

———

- Some libraries will work with researchers to curate datasets before they are deposited
- They check to see if the files are properly documented, if there is a readme file, and that everything is accessible
  - Ex: Data Curation Network

We provide data infrastructure and tools

___

- Libraries often support memberships for data repositories
    - Ex: at UCSF we facilitate access to Dryad
- Other libraries build in-house data repositories
    - Ex: Deep Blue Data at the University of Michigan
- Some develop tools for packaging data
    - Ex: ReproZip at New York University

We contribute to data policy

———

- Librarians contribute to national data policies
  - Ex: [MLA and AAHSL respond to NIH Data Management Policy Draft](#)
- Some librarians partner to create local data policies
  - Ex: Ruth Lilly Medical Library at Indiana University School of Medicine

How do we know if it is working?

We evaluate metrics

———

Process

- Are we being invited to present in more classes?
- Are we meeting with more researchers?

Outcomes

- Are data deposits increasing?
- Are more researchers publishing their data alongside their articles?
- Is our research more reproducible?

Our goal is to nudge researchers towards better, more reproducible data publishing practices

Questions?

ariel.deardorff@ucsf.edu