Why good code is not enough; getting the data license right

Jenny Muilenburg | Vicky Steeves | Erik Mitchell

Data dissemination to support reproducibility

"In addition to the common building blocks of reproducibility, data licensing, ownership and sharing policies/tools are fundamental to repro methods"

- 1. How do licenses that researchers sign impact data availability and reuse?
- 2. What restrictions do your universities put on the creation and sharing of data?
- 3. What systems and strategies exist to make data available and how can we use those for the most impact?

Data acquisition program (UCB)



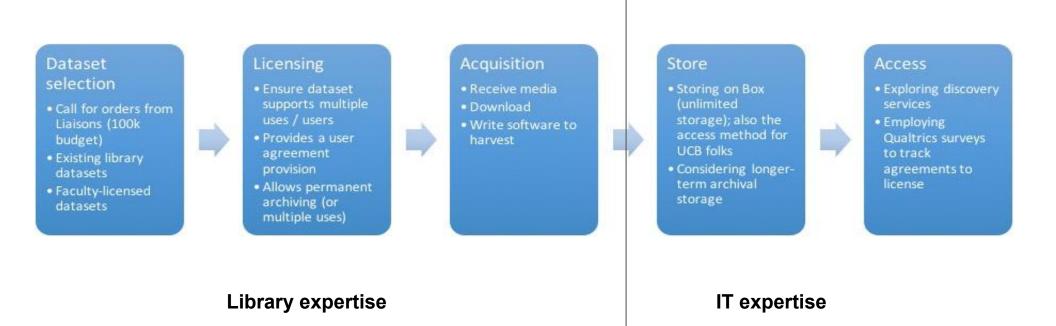
A partnership between the library and D-Lab to acquire and make available licensed or otherwise restricted research data

A framework for licensing, aggregating or otherwise acquiring data leveraging the collection development strength of the library and the storage and auditing strength of the D-Lab

A campus-wide effort that we hope will eventually include connections to HPC, active data storage and RDM

http://guides.lib.berkeley.edu/data

Data acquisition program process



Data ownership

UCB:

Data submitted to UC Dash are released under the terms of the <u>Creative</u> <u>Commons Attribution license</u> (CC-BY-4.0). This means anyone can access the datasets, download them, and use them provided they give proper attribution.

NYU Retention of and Access to Data Policy

https://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/retention-of-and-access-to-research-data.html

UW Research Data Memorandum https://www.washington.edu/research/osp/gim/gim37.html

Data publishing tools for access and reproducibility

How do these tools work together and what impact do they have on enabling or serving as a barrier to access and reproducibility?

Profiles

ORCID - http://orcid.org/0000-0001-8302-1138

GitHub - https://github.com/sr320

Google Scholar - http://scholar.google.com/citations?hl=en&user=aAoYWVUAAAAJ

FigShare - http://figshare.com/authors/Steven_Roberts/97086

ImpactStory -https://impactstory.org/u/0000-0001-8302-1138

Scopus - http://www.scopus.com/authid/detail.url?authorld=55366464800

SlideShare – http://www.slideshare.net/sr320/presentations

Mendeley - http://www.mendeley.com/profiles/steven-roberts1/

Microsoft Academic Search - http://academic.research.microsoft.com/Author/4579019/steven-b-roberts

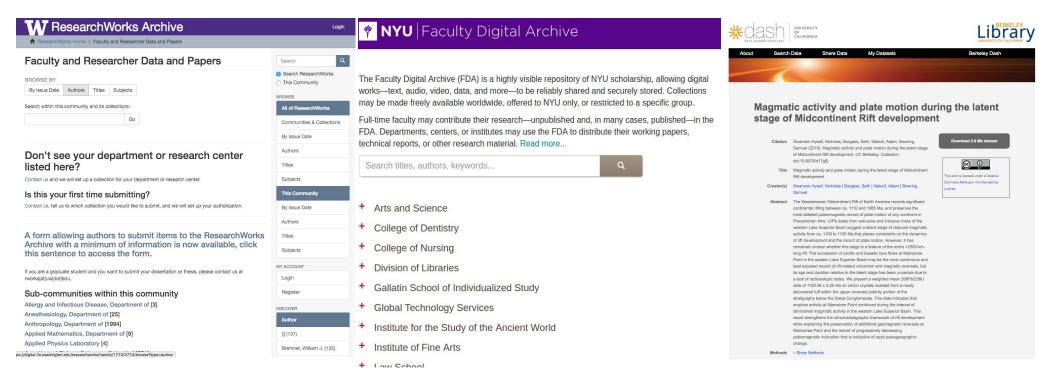
Frontiers - http://community.frontiersin.org/people/StevenRoberts/43641

ResearchGate - http://www.researchgate.net/profile/Steven_Roberts6

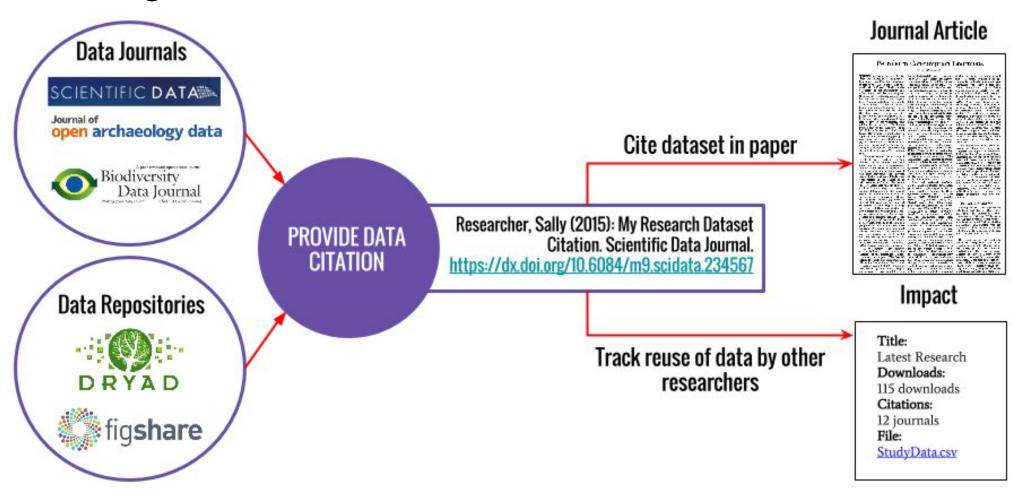
Publons - https://publons.com/author/13038/steven-roberts

From Steven Roberts GitHub page: https://github.com/sr320/LabDocs/blob/master/web/ROS-profile.md

Data publishing solutions @ UW, NYU, UCB



Getting Credit for Your Data



Challenges and strategies

Challenges:

Licensing agreements prohibit data publishing

University agreements mean that researchers cannot publish data

Information systems do not support repro-driven data publishing

Strategies:

Provide programs to shift license agreements to make data publishable

Libraries should help reshape university policies to merge better with data sharing practice

In addition to providing systems; help develop understanding and adoption of identifiers to support data aggregation