

Publishing Data and What it Means for Your Career

Dr. Shelley L. Knuth
Research Computing
University of Colorado-Boulder

http://researchcomputing.github.io/meetup_fall_2014/

What does it mean to publish data?

- Publishing data is sharing data but in a more formal process
- Publishing data can mean lots of different things to different people
 - Website
 - Data repository
 - Discipline specific
 - General
 - Formal
 - Informal
 - Peer reviewed publication

Why publish data?

- Responding to funding agency requirements
 - NSF, NIH, others are following
- Requirement for submission of journal publications
 - Reproducibility and accountability
- Advance research
- Increase citations
- Establish a record for yourself in your field

What data should be published?

- Any data that is needed to reproduce any scientific findings you have made
 - Data
 - Metadata
 - Software
 - Scripts

Publishing Data

- How long should data be published for?
 - What is the cost of maintaining the data?
 - Is it unique?
 - What is its value to the field? Does it have high re-use?
 - How expensive was it to obtain?
 - What does your funder require?
- What about secure data?
 - Security
 - HIPAA
 - FERPA
 - Intellectual Property

Journal Requirements for Accessible Data

- Nature most well known journal
 - Many other journals – Joint Data Archiving Policy (JDAP)
 - “Authors are required to make materials, data and associated protocols promptly available to readers without undue qualification”
 - Ramifications if data not available after publication
 - Prefer availability through public repositories
 - Promote reproducibility

Journal Requirements for Accessible Data

- Other data policies:
- PLOS One:
<http://www.plosone.org/static/policies#sharing>
- American Geophysical Union (AGU):
- <http://publications.agu.org/author-resource-center/publication-policies/data-policy/>

Where can I publish my data?

- Repositories
 - Good for storing and preserving data
 - Provide dedicated curation and maintenance of data
 - Associate DOI with dataset
- What is a DOI?
 - Digital Object Identifier
 - Persistent and unique identifier of digital entities
 - It's a way of ensuring your data is easily discoverable

Digital Object Identifiers

- What does a DOI look like?

[10.1594/USAP/0739464](https://doi.org/10.1594/USAP/0739464)

Links to the URL:

<http://dx.doi.org/10.1594/USAP/0739464>

- How is a DOI cited?
- Citation: Knuth, S. L., Cassano, J. J., Maslanik, J. A., Herrmann, P. D., Kernebone, P. A., Crocker, R. I., and Logan, N. J.: Unmanned aircraft system measurements of the atmospheric boundary layer over Terra Nova Bay, Antarctica, Earth Syst. Sci. Data, 5, 57-69, doi:10.5194/essd-5-57-2013, 2013.

Digital Object Identifiers

- How can I get a DOI?
 - Generated by specific organizations
 - Publish paper (if registering DOIs)
 - Publish dataset (if registering DOIs)
- Free* option:
- www.figshare.com
- Store your data and get a DOI associated with it

*for public data

Publishing Data - Repositories

- How do I find a repository?
 - Is there a discipline specific repository common in your field?
 - In my previous life: <http://www.usap-data.org/>
 - If you don't know of one, try searching at <http://databib.org/>
 - Or contact data-help@colorado.edu and we will find one for you!

Popular Repositories

- Dryad (www.datadryad.org)
 - Popular in life sciences
 - Must be associated with science/medical published article
 - Integrated with publication
 - Embargo of data until publication
 - Curator checks your data
 - Get a DOI
 - Usage statistics
 - Generally up to 10 GB per article; more if needed
 - Not HIPAA compliant
 - There is a fee for upload (organizations usually pay)
 - Free to download
 - Example:
<http://www.datadryad.org/resource/doi:10.5061/dryad.7gs71>

Popular Repositories

- Figshare (www.figshare.com)
 - Free* place to store your data in the cloud
 - Publicly available
 - Citable – get a DOI (if data is public)
 - Pricing:
 - 1 GB – private storage is free
 - Unlimited data shared publicly is free
 - Other charges range from \$8-15/month
 - Promote data on social media
 - Example:
[http://figshare.com/articles/
The_Relationship_Between_Tree_Sapling_Density_and_Competi
ng_Adult_Density_Between_North_American_Deciduous_Trees/
1228724](http://figshare.com/articles/The_Relationship_Between_Tree_Sapling_Density_and_Competing_Adult_Density_Between_North_American_Deciduous_Trees/1228724)

Popular Repositories

- ICPSR (<https://www.icpsr.umich.edu/>)
 - Social Science repository
 - 500,000 research files in social science
 - Lots of documentation on how to properly deposit data
 - Data is reviewed, documentation enhanced after submission
 - ICPSR citations harvested by Web of Science
 - Get a DOI
 - Can make some accounts for confidential data
 - Institutions can become members and we benefit
 - CU is a member
 - Example:
<https://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/13719?q=boring&searchSource=icpsr-landing>

Popular Repositories

- PANGAEA (www.pangaea.de/about/)
 - Data publisher for earth, life, and environmental science
 - “Aimed at archiving, publishing, and distributing georeferenced data from earth system research”
 - Data is freely available – open access
 - Get a DOI associated with it
 - Data archived as publication supplements
 - Provide data preparation and quality control
 - Free through public funds, but greatly appreciate donations
 - Example:
<http://doi.pangaea.de/10.1594%2FPANGAEA.143197>

Peer-Reviewed Data Journals

- Very appealing to academics who are judged on publications
- Submit an article that describes dataset, collection methods, etc
- Example:
<http://www.earth-syst-sci-data.net/5/57/2013/essd-5-57-2013.pdf>
- Goes through peer review
- Many require data submission into repositories before publication

Peer-Reviewed Data Journals

- A list of journals to explore:
 - Scientific Data: (<http://www.nature.com/sdata/>)
 - Geoscience Data Journal (<http://www.geosciencedata.com>)
 - Earth System Science Data
(<http://earth-system-science-data.net>)
 - Ecological Archives - supplemental materials to Ecological Society of America journals
(<http://esapubs.org/archive/default.htm>)
 - Journal of Open Psychology Data
(<http://openpsychologydata.metajnl.com/>)

- Many others listed here:
<http://proj.badc.rl.ac.uk/preparde/blog/DataJournalsList>

Contact Information

- CU Boulder has many services available to you free of charge
- Research Data Services
 - data.colorado.edu
 - data-help@colorado.edu
 - Twitter: @cu_data
 - Facebook: CU Boulder Data
- My personal contact information:
 - shelley.knuth@colorado.edu
 - Twitter: @shelley_knuth