How to assist researchers in sharing their research data | October 22, 2015

Lisa R. Johnston
Research Data Management/Curation Lead and Co-Director of the University Digital Conservancy
University of Minnesota

Alex Ball
Research Data Librarian
University of Bath

Joe Shell
Head of Research Data Management
Mendeley
Data Sharing

Five ways that YOUR Library can support researchers when sharing their data

Library Connect Webinar:
How to assist researchers in sharing their research data

Lisa Johnston
University of Minnesota - Twin Cities
October 22, 2015
5 ways that your Library might help researchers share their data:

1. Keep doing what you do….be an information resource
2. Educate on data management skills/best practices
3. Develop policy + institutional guidelines for data
4. Create a data sharing service: Lots of options!
5. Curate and archive institutional data for reuse
1. Keep doing what you do….be an information resource
MANAGING YOUR DATA

Got data? We’re here to help you manage, share, and preserve your research data. In addition to our Data Repository for the U of M curation services, the Libraries will help you navigate available campus resources throughout the data lifecycle:

Before Your Research Begins

- Schedule a data management plan (DMP) consultation (Request Form) or use our DMP templates
- Explore funding agency requirements for data and learn best practices for getting IRB approval for sharing data.
- See more tools for planning for data management

During Your Research

- Attend workshops and explore online training resources on best practices for data

http://lib.umn.edu/datamanagement
Bring data service providers together in an informal way

Past RDM Discussion Topics:

- Data Storage Options on Campus
- Metadata Standards
- Spatial Data
- Best Practices for De-identifying Research Data
- Data Repositories (Local, National)
- Data Services at the Supercomputing institute
- Practical Examples for Managing data (Sciences)

https://sites.google.com/a/umn.edu/rdm-cop/home
Keep up-to-date information for administration

FUNDING AGENCY GUIDELINES

The 2013 Public Access to Federally Funded Research memo from the White House Office of Science and Technology Policy directed most grant-funding agencies to develop policy requirements public access to resulting articles and data. This page looks at the requirements from federal funders for managing and sharing research data; for articles, see requirements for public access to publications.

The Libraries can help you write and implement your data management plan to meet funder requirements. You can also use our Data Repository for the University of Minnesota (DRUM) to meet federal funder requirements for data sharing.

Federal funders’ responses to the OSTP memo (updated as they become available):

<table>
<thead>
<tr>
<th>Agency</th>
<th>Directorate</th>
<th>Effective Date</th>
<th>DMP Required?</th>
<th>Sharing Venues</th>
</tr>
</thead>
<tbody>
<tr>
<td>DOD</td>
<td>November 2015</td>
<td>End of 2016</td>
<td>Yes</td>
<td><em>Established, publicly accessible institutional repositories</em></td>
</tr>
<tr>
<td>DOE</td>
<td>October 1, 2014</td>
<td>October 1, 2015</td>
<td>Yes</td>
<td>Some data to Open Energy Information Platform; other specified</td>
</tr>
<tr>
<td>HHS</td>
<td>November 2015</td>
<td>November 2015</td>
<td>Yes</td>
<td>Yes, a commercial repository</td>
</tr>
<tr>
<td>ASPR</td>
<td>October 1, 2014</td>
<td>October 1, 2014</td>
<td>Yes</td>
<td>Yes, a publicly accessible database</td>
</tr>
<tr>
<td>CDC</td>
<td>June 2015</td>
<td>October 1, 2015</td>
<td>Yes</td>
<td>Office of Biomedical Informatics, the Office of Science, Technology, and Resources (OSITR)</td>
</tr>
<tr>
<td>FDA</td>
<td>October 1, 2015</td>
<td>October 1, 2015</td>
<td>Yes</td>
<td>Publicly accessible database, specific databases</td>
</tr>
<tr>
<td>NIH</td>
<td>November 2015</td>
<td>November 2015</td>
<td>Yes</td>
<td>Existing, publicly accessible repositories</td>
</tr>
</tbody>
</table>

More Information:
- [http://lib.umn.edu/datamanagement/funding](http://lib.umn.edu/datamanagement/funding)
2. Educate on data management skills/best practices
Offer workshop on “How to write a Data Management Plan (DMP)”

- For researchers
- Discussion Based
- RCR CE Credit
- Departments request custom sessions
- Co-teach with Liaison


https://www.lib.umn.edu/datamanagement/workshops
Provide DMP Templates and offer in-person consultations

- One-on-One consults
- DMP Template
- Boilerplate text
- DMPOnline Tool (CDL)

Data Management Plan

V1 last updated MM-DD-YYYY

<table>
<thead>
<tr>
<th>Name of student/researcher(s)</th>
<th>Your Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of group/project</td>
<td>Project Name or Research Lab (for group plan)</td>
</tr>
<tr>
<td>Funding body(ies)</td>
<td></td>
</tr>
<tr>
<td>Partner organisations</td>
<td></td>
</tr>
<tr>
<td>Project Duration</td>
<td>Start: MM-DD-YYYY  End: MM-DD-YYYY</td>
</tr>
<tr>
<td>Date Written</td>
<td>MM-DD-YYYY</td>
</tr>
</tbody>
</table>

Table of Contents

1. Introduction
2. Data Types
   Section 2 Checklist
3. Data Organization, Documentation and Metadata
   Section 3 Checklist
4. Data Access and Intellectual Property
   Section 4 Checklist
5. Data Sharing and Reuse
   Section 5 Checklist
6. Data Preservation and Archiving
   Boilerplate text: Section 6 Checklist

1. Introduction

Note: Your DMP for NSF grants should not exceed two pages. Contact Research Services in the Libraries for consultation (johndoe@umn.edu).

https://www.lib.umn.edu/datamanagement/DMP
Graduate programs do not always include data information literacy (DIL) skills/competencies.
U of MN Data Management Online Course

- Hybrid online and in person workshops
- Structured around DMP
- Hands on activities
- Direct application to their data

Session 5: How to Digitally Preserve Your Data for the Future

What happens to your data after the project is complete? Will you be able to use the data 10 years from now? Who is going to maintain the data for future use? This module will introduce preservation and curation techniques used by information professionals who manage digital information for long-term access. After this module you will be able to:

- explain the lifespan of potential use for your data in order to recognize the long-term value of your data, and
- identify the relevant preservation-friendly file format for your research data in order to ensure long-term access to your digital information.


http://z.umn.edu/teachdatamgmt
Scaffold DIL skills for undergrads using Personal Information management tools/stories

Archiving Your Photos

Our phones, digital cameras, tablets, and computers all contain the digital record of our lives. How do we keep track of our vast photo libraries? Here are two solutions to get you started.

A Savvy Solution Using Dropbox

Engineering Librarian, Jan Fransen, shares her method of archiving photos using cloud based storage and Wi-fi. View this Diagram of her process and watch the video tutorial.

Which Photo Software Should I Choose?

Use this Matrix of Photo Software tools, such as Flicker, iPhoto, Picasa, and Facebook, to see which one meets your needs. Contact Scott Spitzer with photo management questions.
3. Develop policy + institutional guidelines for data
Institution-wide data policies define roles and responsibilities for long-term data management issues

Also read:

Policy: http://policy.umn.edu/research/researchdata
Ensures accessibility and preservation of research data through curation, metadata, repositories, and other access and retrieval mechanisms to meet federal, state, sponsor, and University requirements.

Trains and supports researchers in the creation and implementation of data management plans.
4. Create a data sharing service: Lots of options!
wood sawdust extractable in some cases with the drill bit showed a biasing effect on DAS, an important consideration in sample weight requirements that corroborates the 100–101 mg sample size used by Shortle et al. [40] (Fig. 5). If coupled with L:D, sample size requirements would be limited by Klason lignin needs, likely best with at least 1 g of field material (fresh wt). Collectively, the results from this trial reinforce the decay class II/III target from the lab trials (given proper field identification when sampling logs) and it demonstrates how a prelimi-
## Data Sharing Techniques

<table>
<thead>
<tr>
<th>Ways to Share your Data</th>
<th>Pros?</th>
<th>Cons?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post online to a personal or project website</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Publish data in a journal as a “supplement” to your main research article.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Make your data “Available on request” via email or dropbox to those who ask.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deposit in a disciplinary repository (e.g. Dryad, FlyBase, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deposit in a general/commercial repository (e.g. FigShare, Mendeley, etc.)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deposit in an institutional repository, such as the Data Repository for the University of Minnesota (DRUM)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Lisa Johnston, University of Minnesota Libraries (ljohnst0@umn.edu)
Available to U of M researchers and provides:

- Open access
- Curation services
- Permanent identifiers (DOI)
- Flexible Licenses
- File download analytics
- Preservation
Data and visualizations of air quality impacts of conventional and alternative light-duty transportation in the United States


Abstract
This is the supporting information for an article entitled "Life cycle air quality impacts of conventional and alternative light-duty transportation in the United States", published in the Proceedings of the National Academy of Sciences, United States (dx.doi.org/10.1073/pnas.1406863111). The study assesses the life cycle air quality impacts of health of 15 alternatives to conventional gasoline vehicles, including vehicles powered by diesel, natural gas and electricity. This supporting information is comprised of 1) a Microsoft Excel file containing emissions and disaggregated by life cycle stage for each scenario; 2) maps of ground-level concentrations of 13 different air pollutants attributable to each scenario; and 3) tables showing temporal variation in ground-level fine particulate matter and ozone (O3) concentrations attributable to each scenario. The data were generated using state-of-the-art pollutant emission and transport modeling.

Creative Commons License: Attribution-NonCommercial-ShareAlike 3.0 United States

Suggested Citation

Usage Statistics for Item: Data and visualizations of air quality impacts of conventional and alternative light-duty transportation in the United States

Total Downloads (2014-Present)

<table>
<thead>
<tr>
<th>File Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>573</td>
</tr>
</tbody>
</table>

Total Monthly Downloads

<table>
<thead>
<tr>
<th>File Downloads</th>
</tr>
</thead>
<tbody>
<tr>
<td>October, 2015</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>September, 2015</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>August, 2015</td>
</tr>
<tr>
<td>24</td>
</tr>
<tr>
<td>July, 2015</td>
</tr>
<tr>
<td>23</td>
</tr>
<tr>
<td>June, 2015</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>May, 2015</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>April, 2015</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>March, 2015</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>February, 2015</td>
</tr>
<tr>
<td>87</td>
</tr>
<tr>
<td>January, 2015</td>
</tr>
<tr>
<td>119</td>
</tr>
<tr>
<td>December, 2014</td>
</tr>
<tr>
<td>172</td>
</tr>
<tr>
<td>November, 2014</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

By using these files, users agree to the Terms of Use. Content distributed via the University of Minnesota's Digital Conservancy may be subject to additional license and use restrictions applied by the depositor.
Data Repository of the University of Minnesota (DRUM)

- Utilize existing repository technologies for cost savings/efficiencies (DSpace, open source software)

Custom upload form and metadata schema for research data

Apply Creative Commons licenses

Curation workflow allows for review of data before openly available
5. Curate institutional research data for sharing and long-term preservation/reuse
What is data curation?

*Data curation* steps may include appraisal, ingest, arrangement and description, metadata creation, format transformation, dissemination and access, archiving, and preservation of *digital research data.*

Twin Cities Housing GIS Data, UMN
What was the process to curate the data?

<table>
<thead>
<tr>
<th>Stage 0:</th>
<th>Stage 1:</th>
<th>Stage 2:</th>
<th>Stage 3:</th>
<th>Stage 4:</th>
<th>Stage 5:</th>
<th>Stage N:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Receive Data</td>
<td>Appraise / Inventory</td>
<td>Organize</td>
<td>Treatment Actions / Processing</td>
<td>Description / Metadata</td>
<td>Access</td>
<td>Reuse Data</td>
</tr>
</tbody>
</table>

Workflow Stages drafted by the “Digital Curation Sandbox” participants borrowed from DCC Curation Lifecycle.

http://hdl.handle.net/11299/162338
What happens after submission to DRUM?

After submission, U of Minn researcher receives a confirmation email.

Within two business days, we will review their data and contact them about proposed modifications to the submission.

- Missing files
- Changes/additions to data documentation
- Reshaping directory structure
- Converting proprietary software to more archival-friendly formats
Methodological Information

Data collection

Processing

Analysis steps

Data-Specific Information

File abbreviations

Name glossary

This readme.txt file was generated on <YYMMDD> by <Name>

GENERAL INFORMATION

1. Title of Dataset:
2. File Information:
   A. Filename:
   B. Short description:
   C. Filename:
   D. Short description
   E. Filename:
   F. Short description:
   G. If data set includes multiple files related to one another, include relationship here:

3. Principal Investigator Contact Information
   A. Name:
   B. Institution:
   C. Address:
   D. Email:

4. Associate or Co-investigator Contact Information
   A. Name:
   B. Institution:
   C. Address:
   D. Email:

http://z.umn.edu/readme
Data Sharing = Services in Support of Research Data Lifecycle

Grant Prep

Data Management Plan (DMP) Consultation

Data Collection

Metadata Consultation

Project Begins

Data Management

Analysis

Published Results

Preservation

Project Close

Sharing

Data Curation & Repository Services

Data Management Best Practices/Training

Storage
Thanks and questions!

Visit our website
http://lib.umn.edu/datamanagement

Lisa Johnston, DMCI Lead University Libraries, ljohnsto@umn.edu
Making the case for sharing with indicators of research data impact

Alex Ball
University of Bath

Library Connect Webinar:
How to assist researchers in sharing their research data

22 October 2015
Motivation

*Mandates* get things done, but grudgingly. To get high quality data sharing, we need to make it *rewarding*.

We need to demonstrate that datasets have *impact*.

CASRAI Dataset-Level Metrics: http://www.casrai.org/Dataset_Level_Metrics
NISO Altmetrics Initiative: http://www.niso.org/topics/tl/altmetrics_initiative/
The first multi-gene phylogeny of the Macrostomorpha sheds light on the evolution of sexual and asexual reproduction in basal Platyhelminthes

Janssen et al. 2015 MPE dataset dryad
2015. http://doi.org/10.5061/dryad.b5908/1
Is part of http://doi.org/10.5061/dryad.b5908 via DataCite Data

http://mdc.lagotto.io/
PlumX

Publication Process with OA
Publication Year: 2012
Researchers: Jason B. Colditz

<table>
<thead>
<tr>
<th>USAGE</th>
<th>3945</th>
<th>SOCIAL MEDIA</th>
<th>230</th>
</tr>
</thead>
<tbody>
<tr>
<td>Views</td>
<td>3917</td>
<td>Shares</td>
<td>180</td>
</tr>
<tr>
<td>Figshare</td>
<td>3917</td>
<td>Figshare</td>
<td>173</td>
</tr>
<tr>
<td>Downloads</td>
<td>23</td>
<td>Facebook</td>
<td>7</td>
</tr>
<tr>
<td>Figshare</td>
<td>23</td>
<td>Tweets</td>
<td>33</td>
</tr>
<tr>
<td>Clicks</td>
<td>5</td>
<td>figshare.com</td>
<td>33</td>
</tr>
<tr>
<td>Bitty</td>
<td>5</td>
<td>Likes</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Facebook</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>+1s</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Google+</td>
<td>8</td>
</tr>
</tbody>
</table>

http://plu.mx/
Altmetric

http://www.altmetric.com/
ImpactStory

Holly Bik

9 Datasets

[A Research-Driven Data Visualization Framework for High-Throughput Environmental Sequence Data](https://figshare.com/)
(2013) figshare.

[view dataset]

[Data from: Dramatic shifts in benthic microbial eukaryote communities following the Deepwater Horizon oil spill](https://dryad.org/)

[view dataset]

[PhyloSift: Phylogenetic analysis of genomes and metagenomes](https://figshare.com/)
(2013) figshare.

[view dataset]

https://impactstory.org/
Other notable services

9.13
RG Score

A new way to measure scientific reputation.
The RG Score takes all your research and turns it into a source of reputation.

PERCENTILE
Your score is higher than 40% of ResearchGate members'.

ResearchGate – https://www.researchgate.net/
Other notable services

<table>
<thead>
<tr>
<th>Citation indices</th>
<th>All</th>
<th>Since 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citations</td>
<td>421</td>
<td>364</td>
</tr>
<tr>
<td>h-index</td>
<td>11</td>
<td>10</td>
</tr>
<tr>
<td>i10-index</td>
<td>15</td>
<td>12</td>
</tr>
</tbody>
</table>

Google Scholar – https://scholar.google.com/
Other notable services

Microsoft Academic Search – http://academic.research.microsoft.com/
Data papers and data journals

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>key</td>
<td>smiles</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>2-(2,4-dinitrobenzyl)pyridine</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>2-[1-piperidinyl]aniline</td>
</tr>
<tr>
<td>4</td>
<td>3</td>
<td>2-[1-piperazinyl]pyrimidine</td>
</tr>
<tr>
<td>5</td>
<td>4</td>
<td>2-[1-piperazinyl]phenol</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>2-[1-cyclohexenyl]ethylamine</td>
</tr>
<tr>
<td>7</td>
<td>6</td>
<td>2-[1-boc-4-piperidinyl(oxy)-n-methylacetamide]</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>2-[1-boc-4-piperidinyl(oxy)-n-cyclopropylacetamide]</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>2-[1-boc-4-piperidinyl(oxy)-n,n-dimethylacetamide]</td>
</tr>
</tbody>
</table>

Published on 20 May 2014 - 17:20 (GMT)
Files size is 2.12 MB

Categories
- Organic Chemistry
- Chemoinformatics

Authors
Jean-Claude Bradley
Antony Williams
Andrew Lang

Tags
- open data
- melting point
- open notebook science

License (what's this?)
CC-0

http://figshare.com/
Final thoughts

Encourage data sharing at your institution:

- Make it as easy as possible to deposit data
- Provide stable bibliographic information and an identifier – this makes tracking citations easier
- Show depositors numbers of views and downloads
- Make these statistics available to other services via an API

Researchers are more likely to take care over sharing if they can see it makes a difference!
Further information


Tools for research data sharing: Mendeley Data and the Electronic Lab Notebook

Joe Shell, 2015

joseph.shell@mendeley.com

Library Connect Webinar:
How to assist researchers in sharing their research data
We agree that research data is important

- Funders are increasingly mandating that data be shared, open and accessible for all the research they fund. In the UK, EU and US, this has implications for **31-34%** of all research carried out.

- Articles that have linked datasets have a **50-70%** higher chance of being cited!*

---

*Piwowar, H.A., Day, R.S., Fridsma, D.B. “Sharing detailed research data is associated with increased citation rate” (2007) *PLoS ONE*
http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0000308

Our current Data situation - Christine L. Borgman

“...issues are becoming clear: the need for coordination among stakeholders, economic challenges to the sustainability of archives, and misaligned public policies for open access to publications and data. The practice and policy issues on the ground are much less well understood, however. Norms for the acquisition, release, and reuse of data – and the very definition of data – vary widely between research domains, and motivations to share data vary accordingly.... Releasing data offers benefits, but so does controlling data. The workforces required for the stewardship of data are many and varied; they too must be nurtured and sustained. Wise investments must be made in knowledge infrastructures – and soon – if research data are to remain useful for generations to come.”

Librarians can be our Data Stewards

Librarians can be the stewards of helping make happy data.

The gatekeepers of not just influencing best practice, but potentially enforcing quality of data collected at point of collection

Benefits

- Reporting
- Meeting funding body requirements
- Non-duplication
- Value of research data at an institute
- Good archiving
Answer – We need to be there every step of the way

Enabling Research | Doing Research | Sharing Research

- Secure funding
- Establish partnerships
- Manage facilities
- Search, discover, read, review
- Collaborate & network
- Experiment
- Synthesize/Analyze
- Manage Data
- Publish and disseminate
- Commer-cialize
- Promote
- Have impact

Search, discover, read, review
- Discover datasets of third parties (world)
- Re-use third party datasets
- Search / query own datasets (overlaps with “manage data”)

Collaborate
- Share data with colleagues in the lab or broader (unstructured) Experimental workflow (structured collab)

Experiment
- Plan and prepare experiments
- Protocol templates
- Capture data from experiments
- Process raw data to enhance value
- Analyse
- Visualise
- Predict/model

Synthesize and analyze
- Store and preserve data
- Integrate data from disparate sources
- Be able to effectively search within all data of your own institution/lab/group

Manage Data
- Publish research dataset

Publish and disseminate
Questions and Thank You!

Lisa R. Johnston
Research Data Management/Curation Lead and Co-Director of the University Digital Conservancy
University of Minnesota
ljohnsto@umn.edu

Alex Ball
Research Data Librarian
University of Bath
A.J.Ball@bath.ac.uk

Joe Shell
Head of Research Data Management
Mendeley
joseph.shell@mendeley.com, @JosephShell