RUTGERS
University Libraries

Data Management Plans and Best Practices

ASEE Engineering Research Council

Laura Palumbo, PE
Chemistry & Physics Librarian/Science Data Specialist

March 8, 2016

Why Data Management?

- Be more efficient
- Increase visibility
- Make it easy for others to find and use your work
- Meet funding requirements
- Preserve your research for the long term

Adapted from http://libguides.rutgers.edu/datamanagement

Without Data Management…

- **Fire!** http://www.nytimes.com/2002/01/13/us/years-of-data-lost-in-fire-at-university.html
- **Flood!** http://www.the-scientist.com/?articles.view/articleNo/33109/title/NYC-Science-Stunned-by-Sandy/
- **Theft!** http://www.news9.com/Global/story.asp?S=13833909
- **Failure!** http://chronicle.com/blogs/wiredcampus/hazards-of-the-cloud-data-storage-services-crash-sets-back-researchers

These are not a backup:

- Backing up your laptop to *an SD Card in the same laptop* is #notabackup
- Backing up to a hard drive that is 6 inches away from your computer is #notabackup
- Backing up your Gmail to *another Gmail account* is #notabackup
- Backing up your book by copying it to another folder is #notabackup

http://www.hanselman.com/blog/TheComputerBackupRuleOfThree.aspx

Your goals as a researcher

- Organize your data
- Store and backup your data
- Preserve data for the future
- Share data with colleagues

From MIT Libraries website

Good Data Management- Organizing

File naming:

- Keep it short if possible, but descriptive. Names should be easily understood by colleagues
- No special characters, no spaces, lowercase
- Use leading zeros-myfile001.tif, not myfile1.tif
- Use non-proprietary filetypes if possible

Versioning:

- All versions of data need to be clearly identified
- **Be consistent! Documentation is key.**

http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/filenameguidelines.pdf

Good Data Management - Storing

Types of data:

- Raw data, working data, processed data, data for reuse

Where will you keep it?

- Department server? Hard drive of your computer? Paper notebooks?

How will you back it up?

- External hard drive? Flash drive? In a drawer in your office? **Use the 3-2-1 Rule**

Good Data Management- Preserving

Preserving for the long term:

- Subject specific repository- re3data.org
- Institutional Repository
- Non-proprietary and standard filetypes and open code are preferred by repositories

A website, YouTube video, or relational database are not ways to ***preserve*** your data!

Good Data Management- Sharing

How will others gain access to your data?

- Will permission be required? (Necessary for sensitive data.)
- What will happen to the data if the PI leaves the institution? Who controls access?
- How will they find your data? Providing DOIs is a good practice.
- Citing data- gives attribution to work by others, increases visibility and impact of work

Good Data Management- Citing

Why Cite Data? From DataCite:

> "Why is it so important to cite data? Books and journal articles have long benefited from an infrastructure that makes them easy to cite, a key element in the process of research and academic discourse. We believe that you should cite data in just the same way that you can cite other sources of information, such as articles and books.

Data citation can help by:

- enabling easy reuse and verification of data
- allowing the impact of data to be tracked
- creating a scholarly structure that recognises and rewards data producers"

Citing Data

- Good data citation includes a persistent identifier, such as a **DOI- Digital Object Identifier**, **URN- Uniform Resource Name**, or **Handle**. (ICPSR)

- How do you get a DOI? Often when you deposit your work in a repository, a DOI is assigned to each item for you. Repositories use a **DOI Registration Agency**, such as **DataCite** or **CrossRef**

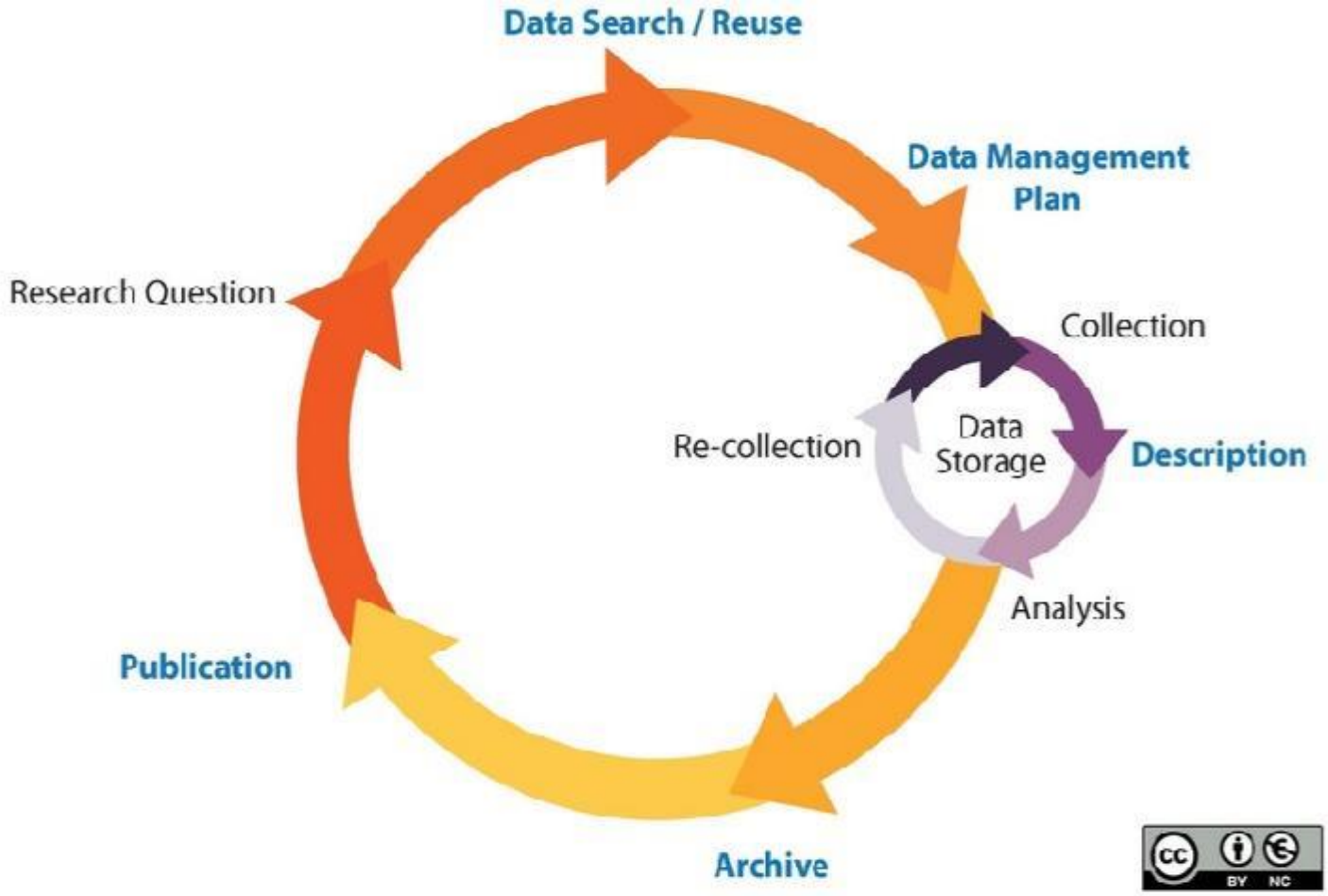
https://www.icpsr.umich.edu/icpsrweb/landing.jsp

Why Share Data?

Image by Pat Marglis,
PLoS Open Access Collection

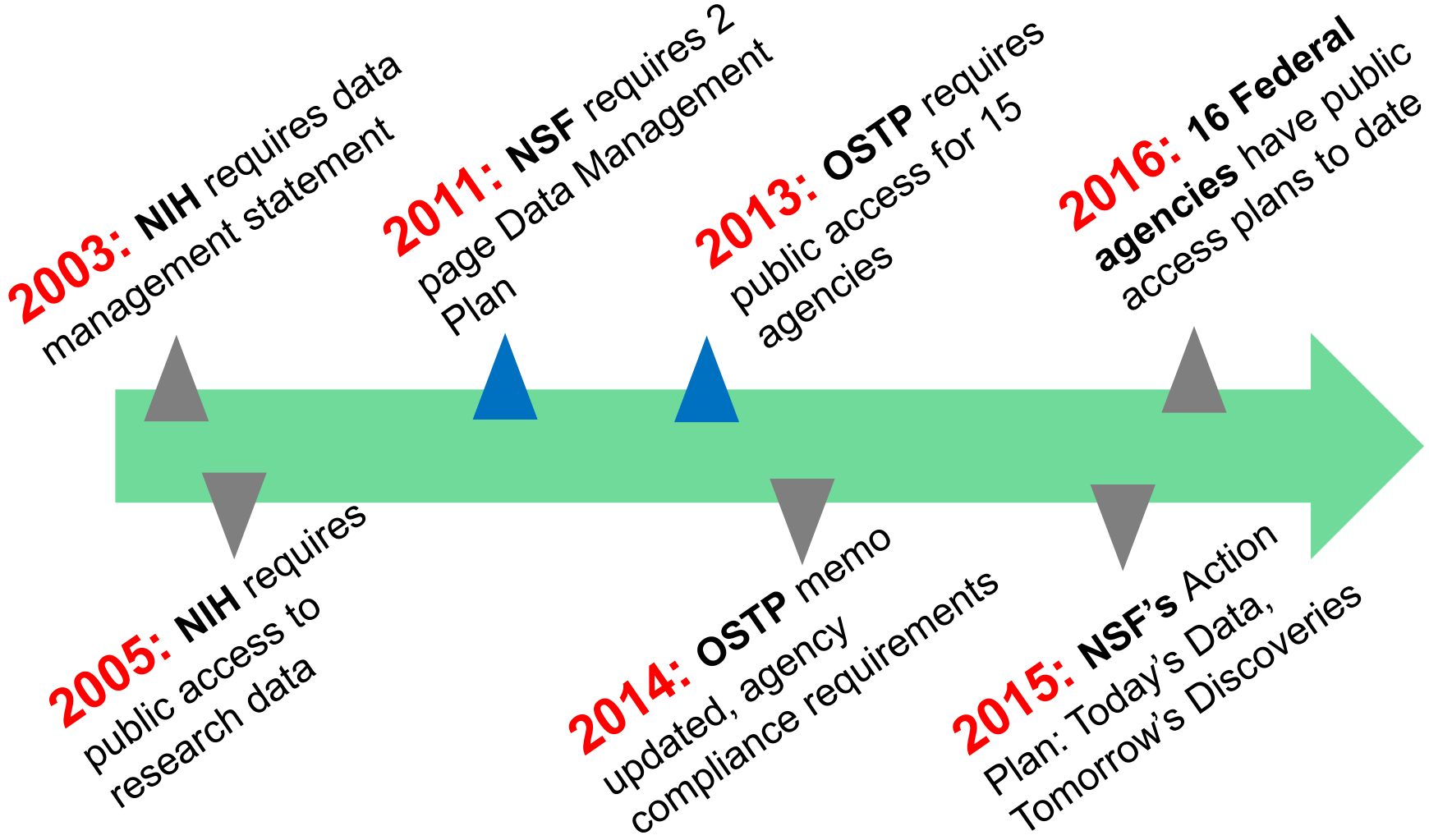
"The purpose of Open Science is not different from that of science itself — open science is simply proper science — reproducible, extensible, accessible." http://opensciencefederation.com/about/

The Research Data Lifecycle



Research Data Lifecycle image from University of California, Santa Cruz
http://guides.library.ucsc.edu/datamanagement/

Federal Funding Requirements Timeline



Federally funded research requires data management

- **NIH** has required a statement concerning data management since 2003 for grants of $500K or more

- As of January 2011, **NSF** is requiring a two page **Data Management Plan**. DMPs are subject to peer review; proposals without a plan will be rejected.
  http://www.nsf.gov/bfa/dias/policy/dmpdocs/che.pdf

- February 2013: Office of Science and Technology Policy requires open access to literature and data for 15 federal agencies (R&D budgets >$100 M)

  http://www.nsf.gov/news/news_summ.jsp?cntn_id=127043

Federal Funding Requirements

- November 2014- OSTP memo is updated, specifying that agencies must state how they expect grantees to comply with sharing requirements.
  https://www.whitehouse.gov/sites/default/files/microsites/ostp/OpenAccess_March-2014.pdf

- March 2015- "NSF's Action Plan: Today's Data, Tomorrow's Discoveries"

- February 2016- 16 Federal agencies have public access plans for sharing research data and publications

Federally funded research will:

- allow a 12 month embargo
- allow full public access to metadata immediately
- allow for exceptions for proprietary or otherwise restricted data
- ensure proper attribution of authors, journals, and original publishers
- ensure that content is preserved and accessible
- require a data management plan from all funded research
- allow researchers to include the cost of complying with the policies in their budgets

Based on Research Data Management Service Group, Cornell University, “Office of Science and Technology Policy Announces Open Access Requirement for Publications and Data”

Federal Requirements for Public Access to Research Data

NIH Data management plans should include:

- **the expected schedule for data sharing**
- **the format of the final dataset (filetype, software used)**
- **the documentation to be provided**
- **any analytic tools that will be provided**
- **if a data-sharing agreement will be required and a brief description of such an agreement**
- **the mode of data sharing (website, repository)**

http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

NSF Data Management Plans

An **NSF DMP consists of 5 components**, which describe:

- **Products of research:** spectra, diffraction patterns, physical properties, computational strategies, software, numerical results, etc. Also quantities- gigabytes, terabytes…
- **Data Format:** instrument output, html, file types- .jpg, .tif... Conversions may be necessary. File names and versions should be standard. **Metadata** is required.
- **Access to Data and Data Sharing Practices and Policies:** how your data will be made freely accessible. Websites, Hydrogen Storage Materials Database, Center for Engineering Strong Motion Data…

DMP components

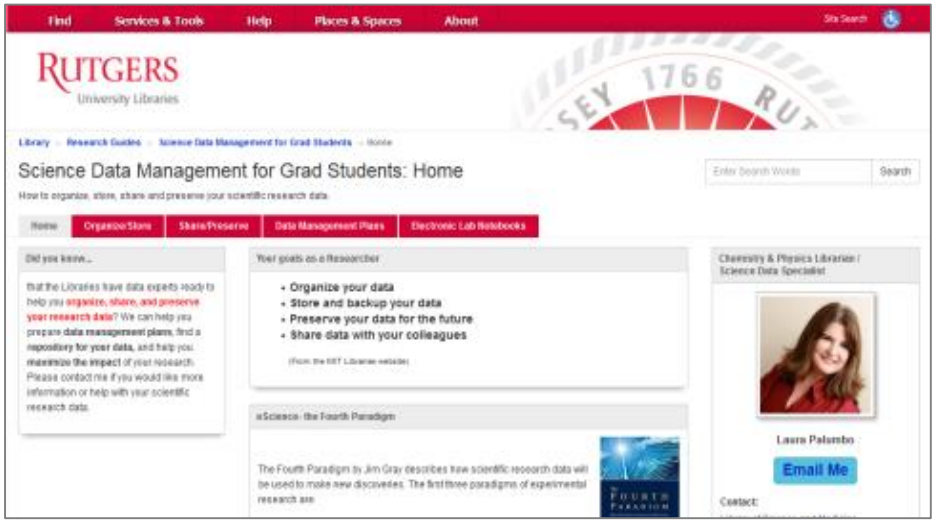
- **Policies for Re-Use, Re-Distribution, and Production of Derivatives:** Who will be able to use your data? Will there be disclaimers on your website? Conditions concerning publication?

- **Archiving of Data:** How will it be preserved? Will hardcopies be transferred to digital format? Will digital media be migrated? Software?

This is a simplified list from NSF- see the Engineering Directorate for more information http://nsf.gov/eng/general/ENG_DMP_Policy.pdf

Resources

Operated by the California Digital Library, the DMPTool is a site supporting general DMP development with some school-specific guidance



Research Guides

offer information about services and experts to support your data management efforts

Adapted from Data Management Best Practices by Womack and Morgan, October 2013

Resources

Digital Curation Centre – Training, DMPonline tool, DMP examples, metadata standards

DataCite – DOI Citation formatter, find repositories and data

re3data.org – Registry of Research Data Repositories

OpenDOAR – Directory of Open Access Repositories

Would someone else be able to find your work?

Register for an ORCID id:

Open **R**esearcher and **C**ontributor **ID**

What is ORCID?

Register for an ORCID id

http://orcid.org

Image from orcid.org

Thank you!

laura.palumbo@rutgers.edu
Twitter: @LauraBPalumbo

References/Resources

DataCite https://www.datacite.org/

DMPOnline. Digital Curation Centre. https://dmponline.dcc.ac.uk/

DMPTool. California Digital Library. https://dmptool.org/

ICPSR. Inter-university Consortium for Political and Social Research. https://www.icpsr.umich.edu/icpsrweb/landing.jsp

National Institutes of Health (NIH) Data Sharing Policy and Implementation Guidance. March 2003. http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#fin

National Science Foundation. (n.d.) Data Management for NSF Engineering Directorate Proposals and Awards. http://nsf.gov/eng/general/ENG_DMP_Policy.pdf

NSF's Public Access Plan: Today's Data, Tomorrow's Discoveries. March 18, 2015. http://www.nsf.gov/pubs/2015/nsf15052/nsf15052.pdf

Office of Science and Technology Policy (OSTP) Memorandum, Increasing Access to the Results of Federally Funded Scientific Research, February 22, 2013.
https://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

Office of Science and Technology Policy (OSTP) Memorandum, Update on Increasing Access to the Results of Federally Funded Scientific Research, March 24, 2014.
https://www.whitehouse.gov/sites/default/files/microsites/ostp/OpenAccess_March-2014.pdf

OpenDOAR- The Directory of Open Access Repositories
http://www.opendoar.org/

Open Science Federation
http://opensciencefederation.com

ORCiD http://orcid.org

Registry of Research Data Repositories
http://www.re3data.org/

Rutgers Libraries Guides on Data Management:
http://libguides.rutgers.edu/grad_sciencedata
http://libguides.rutgers.edu/datamanagement