Compliance: Data Management Plans and Public Access to Data

Margaret E. Henderson
Virginia Commonwealth University

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/escience_symposium

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 4.0 License.
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Compliance: Data Management Plans and Public Access to Data

Hillary Miller
Scholarly Communications Outreach Librarian
hmiller5@vcu.edu
April 6, 2016
NIH Public Access Policy

SEC. 218. The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine’s PubMed Central an electronic version of their final peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication.
Increasing Access to the Results of Federally Funded Scientific Research

February 22, 2013

“ensuring that, ... the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.”

“develop plans to make the results of federally-funded research publically available free of charge within 12 months after original publication.”
Applies to:

Department of Agriculture
Department of Commerce
Department of Defense (DOD)
Department of Education
Department of Energy (DOE)
Department of Health and Human Services
Agency for Healthcare Research and Quality (AHRQ)
Assistant Secretary for Preparedness and Response (ASPR)
Center for Disease Control and Prevention (CDC)
Food and Drug Administration (FDA)
National Institutes of Health (NIH)
Department of Homeland Security
Department of Housing and Urban Development
Department of Interior
Department of Labor
Department of Transportation
Department of Veterans Affairs
Environmental Protection Agency (EPA)
Institute for Museum and Library Services (IMLS)
National Aeronautics and Space Administration (NASA)
National Institute for Standards and Technology (NIST)
National Oceanic and Atmospheric Administration (NOAA)
National Science Foundation (NSF)
Smithsonian Institution (Complying voluntarily. Research budget < $100 million)
Fair Access to Science and Technology Research (FASTR) Act

FASTR calls for federal agencies with extramural research budgets in excess of $100 million to establish consistent, **permanent public access policies for articles** reporting on their funded research. This means that articles reporting on the results of taxpayer-funded research would be made available to the general public to freely access and fully use. Embargo would be no more than 12 months, but potentially could be even shorter.

FASTR would codify the OSTP memorandum to provide greater public access to taxpayer-funded research.

SPARC FASTR resource page
Peter Suber’s FASTR resource page at the Harvard Open Access Project
# Public vs Open Access

<table>
<thead>
<tr>
<th>Public</th>
<th>Open</th>
</tr>
</thead>
<tbody>
<tr>
<td>free of cost to read</td>
<td>free of cost to read</td>
</tr>
<tr>
<td>not free to use or reuse</td>
<td>free to use or reuse, no copyright or licensing restrictions</td>
</tr>
<tr>
<td>usually not final version</td>
<td></td>
</tr>
<tr>
<td>often embargoed</td>
<td>no embargoes</td>
</tr>
<tr>
<td>journal generally owns copyright</td>
<td>author retains copyright</td>
</tr>
</tbody>
</table>
What about data?

• Data management plan and compliance with data sharing

• Public vs open access

• Outreach and education for articles and data will likely happen together
The Research Lifecycle
Data Management

- Write the data management plan.
- Get the funding!
- Do the data management.
When you’re ready to publish....

- Check author’s rights and data policies.
- Negotiate if necessary.
- Write the costs of data storage into the grant.
- Do the same with publishing costs if you want to go open access and there are author fees.
Deposit Your Data

• Deposit data **when**: article is accepted for publication, 30 months after collection, upon publication, within reasonable time period, but NOT when embargo period is over.

• Deposit data **where**: this will depend on the requirements of the funder.

• Link data to the article.
Data Decisions

• Give your data an identifier.
• License your data.
Deposit Your Article

• Deposit the article when the embargo period ends, or

• Let the publisher deposit the article if they offer this service, but

• **MAKE SURE THEY DO IT!**
Wellcome criticises publishers over open access

Wiley and Elsevier attacked over failures to deposit papers as research funder warns it could withdraw funding for hybrid journals

March 24, 2016

By David Matthews  Twitter: @DavidMjourn

The Wellcome Trust has warned big publishers than unless they improve their service and lower their costs it could refuse to provide researchers with funds to publish in certain types of their journals.

https://www.timeshighereducation.com/news/wellcome-criticises-publishers-over-open-access
Why do we have to do this?

• Because we ought to
• Because we should have been doing it all along
• Because we have to
Why? Researchers

• Increases readers’ ability to find and use relevant literature
• Increases the visibility, readership and impact (i.e. citations) of works
• Creates new avenues for discovery in digital environment
• Enhances interdisciplinary research
• Accelerates the pace of research, discovery and innovation
Why? Research Funders

• Leverages return on research investment
• Creates tool/portal to manage research portfolio
• Avoids funding duplicative research
• Creates transparency
• Encourages greater interaction with results of funded research
Why? Educational Institutions

• Contributes to core mission of advancing knowledge
• Democratizes access across all institutions – regardless of size or budget
• Provides previously unattainable access to community colleges, two-year colleges, K-12 and other schools
• Provides access to crucial STEM materials
• Increases competitiveness of academic institutions
• Students
• Enriches the quality of their education
• Ensures access to all that students need to know, rather what they (or their school) can afford
• Contributes to a better-educated workforce

3 why slides from http://www.sparc.arl.org/resources/open-access/why-oa
Retraction Watch

Poll: Should retracted papers be made available for free?

with 5 comments

Recently, Robert Geller of the University of Tokyo brought an interesting issue to our attention. In following a particular paper that had been flagged with concerns on PubPeer, he saw that the journal had eventually retracted it. Even though the journal was sold under a subscription-based model, it made the retraction notice available outside the paywall—per the recommendations of the Committee on Publication Ethics (COPE).

The paper itself now included a link to the retraction notice—also recommended by COPE—but the retracted article remained behind the paywall. In other words, to read the retracted paper, non-subscribers would have to pay $300 Yen (about $30).

Geller contacted us, concerned that the journal was continuing to profit from a retracted paper.

It’s a question we’ve never considered before: Read the rest of this entry.

Now this is transparent: Retraction for plagiarism earns 4–page editor’s note

with one comment

A journal has retracted a paper about 3D imaging after concluding the authors used equations from another researcher without attribution—and has conveniently included a detailed editorial explaining exactly what happened.

It’s rare for us to see a journal be so transparent in explaining what went wrong with one of its papers, so we’re thanking Stuart Cranshaw, from Denbighshire in Wales, UK, the editor of The Photogrammetric Record, for “doing the right thing.”

Even the retraction note is reasonably forthcoming: Read the rest of this entry.

Recent Posts

Weekend reads: Calls for retraction a bad idea? Is scientific fraud a crime?
WE FOSTER THE **OPENNESS, INTEGRITY, AND REPRODUCIBILITY** OF SCIENTIFIC RESEARCH

COS is a non-profit technology company providing free and open services to increase inclusivity and transparency of research. COS supports shifting incentives and practices to align more closely with scientific values.

**What We Work On**

- **Metascience**
  
  COS supports research on scientific practices. These efforts can inform best practices and serve as platforms to demonstrate reproducible research methods. [See some examples.](#)

- **Community**
  
  COS fosters open science communities of researchers, developers, and leaders. Check out [COS Communities](#) to learn more.

- **Infrastructure**
  
  COS supports and maintains the free [Open Science Framework](#) to help researchers manage and archive their research, privately or publicly. [Take a tour](#) to learn more.

**Who We Work With**

- **Scientists**
  
  COS empowers scientists to make their work more accessible and reproducible, and includes

- **Publishers & Societies**
  
  COS maintains free, easy-to-adopt tools and services for journals, societies, and funders to

- **Developers**
  
  COS builds open source web apps, connects via API with other services, and supports open projects.
# Comply with Public Access Mandates

Helpful information about government and other funder mandates for public access to articles and data generated by funded research.

## OSTP Policies from US Federal Agencies

This table covers basic information about the new policies required by the OSTP memo. More information on the memo, and some of the policies, is available on this VCU Libraries page. The agency names link directly to the agency plan.

**US Public Access Policies**

<table>
<thead>
<tr>
<th>Agency</th>
<th>Policy starting date</th>
<th>Article Solution</th>
<th>Embargo Period</th>
<th>Data Management Plan</th>
<th>Data Solution#</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ</td>
<td>Feb 2015 for pubs, Oct 2015 for data</td>
<td>PubMed Central (PMC)</td>
<td>12 months**</td>
<td>Required</td>
<td>Commercial repository, yet to be named</td>
</tr>
<tr>
<td>ALC/NIDILRR</td>
<td>Oct 2016 for pubs, Oct 2017 for data</td>
<td>PMC</td>
<td>12 months</td>
<td>Required</td>
<td>&quot;Publicly accessible databases&quot;</td>
</tr>
<tr>
<td>ASPR**</td>
<td>Oct 2015</td>
<td>PMC</td>
<td>12 months**</td>
<td>Required</td>
<td>Scientific data repositories, data.gov data registry</td>
</tr>
<tr>
<td>CDC*</td>
<td>Oct 2015 for pubs</td>
<td>CDC Stacks, via NIHMS submission system</td>
<td>12 months**</td>
<td>Multiple solutions + data registry</td>
<td></td>
</tr>
<tr>
<td>DOD</td>
<td>Starting FY 2015</td>
<td>Defense Technical Information Center (DTIC)</td>
<td>12 months</td>
<td>Required</td>
<td>No specific solution - decentralized approach</td>
</tr>
<tr>
<td>DOE</td>
<td>Office of Science has already started, others to follow</td>
<td>FAGES (FAQ)</td>
<td>12 months</td>
<td>Required - merits evaluated. Further information.</td>
<td>Varies by office</td>
</tr>
<tr>
<td>DOT</td>
<td>2016</td>
<td>National</td>
<td>12-18 months</td>
<td>Required</td>
<td>Conformant repositories</td>
</tr>
</tbody>
</table>
We help you comply with federal public access mandates

The Office of Science and Technology Policy (OSTP) requires that federal agencies with more than $100 million in research and development expenditures make the results of that research freely available to the public within one year of publication. Researchers who receive these funds must also account for, manage and share the digital data resulting from this publicly funded scientific research.

What should you do to comply with federal public access policies?

- Check for funders public access policies when you apply for grants.
- Sign only agreements with publishers that allow you to comply with your funder's public access policy.
- Retain a copy of your final, peer-reviewed manuscript, the version that incorporates peer review changes but not publisher formatting. You may need to deposit it in an approved repository.
- Add your final, peer-reviewed manuscript to Scholars Compass, VCU's institutional repository, so you will have an easy way to provide a stable link and metadata for your work.
- Make data that supports tables and figures in your publications publicly accessible. Scholars Compass can hold your data and link to your article.

What federal agencies are making changes?

- Agency for Healthcare Research & Quality (AHRQ)
- Assistant Secretary for Preparedness and Response (ASPR)
- Centers for Disease Control (CDC)
- Department of Commerce (DOC)
- Department of Homeland Security (DHS)
- Department of Defense (DOD)
- Department of the Interior (DOI)
- Department of Education (DOE)
- Department of Transportation (DOT)
- Environmental Protection Agency (EPA)
- Food and Drug Administration (FDA)
- National Aeronautics and Space Administration (NASA)
- National Institutes of Health (NIH)
- National Institute of Standards and Technology (NIST)
- National Oceanic and Atmospheric Administration (NOAA)
- National Science Foundation (NSF)
- Office of the Director of National Intelligence (ODNI)
- United States Department of Agriculture (USDA)
- United States Agency for International Development (USAID)
- U.S. Geological Survey (USGS)
- Veterans Administration (VA)

Learn more at:
guides.library.vcu.edu/publicaccess

Margaret Henderson
(804) 828-2714
mehenderson@vcu.edu

Hillary Miller
(804) 827-0462
hmiller54@vcu.edu
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH</td>
<td><a href="http://grants.nih.gov/grants/guide/notice-files/NOT-">http://grants.nih.gov/grants/guide/notice-files/NOT-</a> HS-16-008.html</td>
<td>Feb 2015 (A)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), with article publication (D)</td>
<td>partial</td>
<td>full</td>
<td>partial</td>
<td>none</td>
<td>full</td>
<td>full</td>
<td>PubMed Central</td>
<td></td>
</tr>
<tr>
<td>ASPR</td>
<td><a href="http://www.osdr.gov/PreparationPlanning/sciencePages/AccessPlan.pdf">http://www.osdr.gov/PreparationPlanning/sciencePages/AccessPlan.pdf</a></td>
<td>Oct 2015 (A, D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), with article publication/within 30 months of collection (D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>full</td>
<td>full</td>
<td>PubMed Central</td>
<td></td>
</tr>
<tr>
<td>CDC</td>
<td><a href="http://www.cdc.gov/oia/oia/policy/2016/CDC-Public-Access-Plan_Jun2015">http://www.cdc.gov/oia/oia/policy/2016/CDC-Public-Access-Plan_Jun2015</a> phenomenal.pdf</td>
<td>Jul 2013 (A), Oct 2015 (D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), with article publication/within 30 months of collection (D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>full</td>
<td>full</td>
<td>CDC Stacks</td>
<td></td>
</tr>
<tr>
<td>DOD</td>
<td><a href="http://www.defense.gov/dodforms/DOD_Public-Access-Plan_Feb2015.pdf">http://www.defense.gov/dodforms/DOD_Public-Access-Plan_Feb2015.pdf</a></td>
<td>estimate fiscal year 2015</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), within a reasonable time (D)</td>
<td>full</td>
<td>full</td>
<td>none</td>
<td>none</td>
<td>full</td>
<td>full</td>
<td>Defense Technical Information Center</td>
<td></td>
</tr>
<tr>
<td>DOE</td>
<td><a href="http://www.energy.gov/democracy/smart-policy-digital-digital-research-data-management">http://www.energy.gov/democracy/smart-policy-digital-digital-research-data-management</a></td>
<td>Oct 2014 (A), Oct 2014 (D - Office of Sciences), Oct 2015 (D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), with article publication (D), data sharing policy will be evaluated &quot;beginning about three years after this policy goes into effect&quot;, or Oct 2018</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>full</td>
<td>full</td>
<td>PAGES to index, article hosts choices: [list publisher, 2nd local repository, 3rd OSTI CD collect all accepted manuscripts]</td>
<td></td>
</tr>
<tr>
<td>DOT</td>
<td><a href="https://www.transportation.gov/openofficial.dot-public-access-plan">https://www.transportation.gov/openofficial.dot-public-access-plan</a></td>
<td>Jan 2016 (1st report on implementation)</td>
<td>full</td>
<td>TBD</td>
<td>partial</td>
<td>Within 12 months (A), not specified, ASAP (D)</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>none</td>
<td>TBD</td>
<td>full</td>
<td>National Transportation Library (NTL)</td>
<td></td>
</tr>
<tr>
<td>FDA</td>
<td><a href="http://www.fda.gov/downloads/ScienceResearch/AboutScienceResearch">http://www.fda.gov/downloads/ScienceResearch/AboutScienceResearch</a></td>
<td>Oct 2015 (A, D)</td>
<td>full</td>
<td>full</td>
<td>full</td>
<td>Within 12 months (A), with article publication (D)</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>partial</td>
<td>full</td>
<td>full</td>
<td>PubMed Central</td>
<td></td>
</tr>
<tr>
<td>IES</td>
<td><a href="https://ies.ed.gov/funding/researchaccess.asp">https://ies.ed.gov/funding/researchaccess.asp</a></td>
<td>In effect (A, D), FY 2016 (D)</td>
<td>full</td>
<td>partial</td>
<td>full</td>
<td>Within 12 months (A), timely, with article publication (D)</td>
<td>full</td>
<td>full</td>
<td>partial</td>
<td>partial</td>
<td>full</td>
<td>full</td>
<td>ERIC</td>
<td></td>
</tr>
</tbody>
</table>

Questions?

Margaret Henderson
Director, Research Data Management
VCU Libraries
(804) 628-2714
mehenderson@vcu.edu

Hillary Miller
Scholarly Communications Outreach Librarian
VCU Libraries
(804) 827-0462
hmiller5@vcu.edu
Compliance: Data Management Plans and Public Access to Data

Margaret Henderson
Director, Research Data Management
mehenderson@vcu.edu
April 2016
Real scientist

We need you to replicate this top-secret research.

REPLICATE?!
I can barely understand the writing! The graphs have no legends, there's raw data pasted everywhere, and it has no experimental context whatsoever!

Movie scientist

We need you to replicate this top-secret research.

REPLICATE?!
Give me an hour and I can do it better.

http://creativecommons.org/licenses/by-nc-nd/3.0/  http://theupturnedmicroscope.com/
FEDERAL REQUIREMENTS
NIH Data Sharing Policy

“Data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidential and proprietary data. To facilitate data sharing, investigators submitting a research application requesting $500,000 or more of direct costs in any single year to NIH on or after October 1, 2003 are expected to include a plan for sharing final research data for research purposes, or state why data sharing is not possible.“

NIH Genomic Data Sharing Policy

“Policy for Sharing of Data Obtained in NIH Supported or Conducted Genome-Wide Association Studies (GWAS) (effective January 2015)

• “For the purposes of this policy, a genome-wide association study is defined as any study of genetic variation across the entire human genome that is designed to identify genetic associations with observable traits (such as blood pressure or weight), or the presence or absence of a disease or condition.”
• Applies to all NIH-funded research that generates large-scale human or non-human genomic data, as well as the use of those data for subsequent research.
• Requires “Genomic Data Sharing Plan”.
• Allows for expenses in project budget.
• Requires public availability of data in a “timely manner.”
• Recommends NIH-funded or third-party repositories for deposition.

NSF Policies

NSF Data Sharing Policy
Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4. http://www.nsf.gov/bfa/dias/policy/dmp.jsp

NSF Data Management Plan Requirements
Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled “Data Management Plan”. This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation. https://www.nsf.gov/eng/general/dmp.jsp
Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.


NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan." This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

https://www.nsf.gov/eng/general/dmp.jsp

Slide courtesy of Amanda Whitmire
Increasing Access to the Results of Federally Funded Scientific Research - February 22, 2013

“ensuring that, ... the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.”

“develop plans to make the results of federally-funded research publically available free of charge within 12 months after original publication.”

https://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research
THE PLANS THREE YEARS ON

http://guides.library.vcu.edu/publicaccess
Data Management Plans

• All agencies will require a data management plan.
• “Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.” DOE third principle [http://science.energy.gov/funding-opportunities/digital-data-management/](http://science.energy.gov/funding-opportunities/digital-data-management/)

• DOE and NSF have indicated they will review and evaluate DMPs
Data Sharing

• Digitally formatted data arising from unclassified, publicly releasable research and programs.
• Decentralized approach to data storage.
• Allow for inclusion of costs for data management and access.
• Will establish a system to enable the identification, attribution, (federated) storage, and access of digital data.

From NASA FAQ
• “First of all, be reassured that we are not going to force you to reveal your precious proprietary data prior to publication. No personal, proprietary or ITAR data is included.”

http://science.nasa.gov/researchers/sara/faqs/dmp-faq-roses/
Guiding Principles and Common Approach for Enhancing Public Access to the Results of Research Funded by HHS Operating Divisions (February 2015)

Provides continuity across Operating Divisions: NIH, CDC, FDA, AHRQ, and ASPR (voluntary)

- HHS will take a common, stepped approach to establishing data policy and infrastructure: assessment, inventory, DMPs, pilots, training
NIH – National Institutes of Health

Data
• Unclassified digital scientific research data
• Submission of DMP
• Deposit of data into appropriate, existing, publicly accessible repositories, including NIH data repositories
• Upon acceptance for publication (will explore)
• Include appropriate costs in proposals
• Utilize existing reporting structures
• “enforcement actions” including withholding of funds

Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (February 2015)
CDC – Centers for Disease Control and Prevention

Data

• Unclassified digital scientific research data
• Submission of DMP (generic CDC template)
• Deposit of data into suitable platform (tbd)
• Upon acceptance for publication or within 30 months of collection
• Include appropriate costs in proposals
• Utilize existing reporting structures
• Reduction or restriction of funds, award termination, negative influence on future awards
FDA – Food and Drug Administration

Data
• Digitally formatted scientific data resulting from unclassified research
• Submission of DMP
• Deposit of data into discipline-specific repositories
• Upon acceptance for publication
• Include appropriate costs in proposals
• Utilize existing reporting structures
• Termination of contract or grant; withholding of funds

Plan to Increase Access to Results of FDA-Funded Scientific Research (February 2015)
AHRQ – Agency for Healthcare Research and Quality

“To the extent feasible and consistent with applicable law and policy; Agency mission; resource constraints; U.S. national, homeland, and economic security; and the objectives listed below, digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze.”

Data
• Unclassified research data (check Scope section)
• Submission of DMP
• Deposit of data to AHRQ or other repository
• “timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset Include appropriate costs in proposals and applications”
• Utilize existing reporting structures
• Negative influence on future funding

AHRQ Public Access to Federally Funded Research (February 2015)
AHRQ Public Access to AHRQ-Funded Scientific Publications (February 19, 2016)
ASPR – Assistant Secretary for Preparedness and Response

Data

• Digital scientific data
• Submission of DMP
• Deposit of data to recognized scientific repository
• 30 months from creation of data set, or upon publication
• Appropriate costs included in proposals and applications
• Staff and peer review; existing reporting structures
• Negative influence on future funding

Public Access to Federally Funded Research: Publications and Data (February 2015)
DOD

• Digitally formatted data arising from unclassified, publicly releasable research and programs.
• Decentralized approach to data storage.
• Require the submission of data management plans.
• Allows for inclusion of costs in proposals
• Requirements: DMPs cover data sharing and data are available before making subsequent awards
• **DTIC** (Defense Technical Information Center): repository for full-text of peer-reviewed author final manuscripts or publisher versions of research articles and repository for metadata of digital scientific data sets.
• **DOD Public Access search up and running.**
• **Plan**
DOE

• **Statement on Digital Data Management**
• DMP required and merits evaluated.
• Covers unclassified and unrestricted digital research data, i.e. digital data required to validate findings.
• Enterprise Data Inventory -> Public Data Listing -> populates data.gov
• “Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.”
NSF

• All proposals need 2 page DMP that will be part of merit review and be monitored.
• Data in appropriate repository, metadata required.
• Funds available to prepare data for sharing.

Do I have to deposit the data that support findings in my article in a public access repository?

Mandatory deposit of data on which an article is based may be required by the journal publisher or other funders. Data collected as part of NSF-funded research, whether or not they are used to support a given publication, should be managed according to the data management plan.

At a minimum required DMP must promise to release the data needed to reproduce figures, tables and other representations in publications, at time of publication or within reasonable time period.

Publication should provide link to data.

Only the data used to support, validate, and corroborate published research findings are required to be shared, per this plan. Preliminary data, trial data, etc. are not included.

NASA will develop a data catalog

Final plan
DMP FAQ ROSES (Research Opportunities in Space and Earth Science)

“First of all, be reassured that we are not going to force you to reveal your precious proprietary data prior to publication. No personal, proprietary or ITAR data is included.”
ARE YOU DONE YET?
WHAT DOES THIS MEAN FOR YOU?
WHAT IS DATA?

ASK THE DATASAURUS!

meinegenerator.net
Digital Scientific Data: Consistent with the OSTP memorandum and OMB Circular A-110, digital scientific data are defined for the purpose of this plan as “the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.” NIH considers “digital” data to be data that has been recorded in any electronic format that can be accessed using a computer. The definition of digital scientific data includes data that are used to support a scientific publication as well as data from completed studies that might never be published. It may include data that support or refute a hypothesis, but does not include draft or preliminary data sets. For the purpose of this plan, the definition of digital scientific data does not include software per se. NIH recognizes that software and tools such as interview protocols, coding guides, data collection instruments, or manuals may be necessary to access and interpret data. In such cases, and also when the research involves the collection and analysis of administrative data (e.g., data from the Social Security Administration) the data management plan will be expected to address how information about such items will be made available.

59 NIH has policies governing the appropriate sharing of NIH-funded research resources other than data (e.g., model organisms, cell lines, research reagents). Those policies can be found at http://sharing.nih.gov

Digital scientific data that are covered by this policy include:

- Field data.
- Lab data.
- Other data (e.g., quality control samples, sample ID data, and instrument calibration data).

Digital scientific, clinical, or institutional data that can be in scope at the discretion of the program or in an appropriate context:

- Models and model-related content, including parameters and outputs, including models of public health emergencies.
- Command files, such as SAS or STATA files created from secondary sources (typically referred to as secondary or outside data).
- De-identified Electronic Health Record (EHR) patient treatment records.
- Non-proprietary records and data collected as part of the National Hospital Preparedness Program.

Digital scientific data that are not in scope for this policy include:

- Personally identifiable data, however, AHRQ will make an attempt to make all research data available to the public by creating de-identified public use data files.
- Proprietary trade data.
- Data related to protecting critical infrastructure.
- Other data whose release is limited by law, regulation, security requirements, or policy.

http://www.ahrq.gov/funding/policies/publicaccess/index.html#2
BUT

NIH and NSF still have sharing policies that cover more than digital data. Need to include specimens, software, etc., in plans for those agencies.
WRITING YOUR DMP
**Who**
Name of data owner or steward who is responsible for data.

**What**
Data types, samples, software, other materials.

**Describe**
Standards, metadata, if applicable. Readme or Data Dictionary

**Share**
Method for sharing or making data public.

**Reuse**
Note any restrictions or licenses for reuse.

**Preserve**
How long and where data will be kept.
Name of data owner or steward who is responsible for data.
What

Data types, samples, software, other materials.
Examples of Research Data

- Documents (text, Word), spreadsheets
- Laboratory notebooks, field notebooks, diaries
- Questionnaires, transcripts, codebooks
- Audiotapes, videotapes
- Photographs, films
- Protein or genetic sequences
- Spectra
- Test responses
- Slides, artifacts, specimens, samples
- Collection of digital objects acquired and generated during the process of research
- Database contents (video, audio, text, images)
- Models, algorithms, scripts
- Contents of an application (input, output, logfiles for analysis software, simulation software, schemas)
- Methodologies and workflows
- Standard operating procedures and protocols
REAL scientist

What? Where are your data files?
What format are they in?
How many log books do you have?
Which shelf?
What about that old cabinet in your office?
Should I get that too? HEY!

Please...
save my work...

MOVIE scientist

Please...
save my work...

Done.

Describe Standards, metadata, if applicable. Readme or Data Dictionary
Describe – Document Your Data

• Readme File in all Folders
• Metadata
• Data Dictionary
Data from: Female competition in chimpanzees

Files in this package

Content in the Dryad Digital Repository is offered "as is." By downloading files, you agree to the Dryad Terms of Service. To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data.

<table>
<thead>
<tr>
<th>Title</th>
<th>Female competition in chimpanzees</th>
</tr>
</thead>
<tbody>
<tr>
<td>Downloaded</td>
<td>85 times</td>
</tr>
<tr>
<td>Description</td>
<td>UTM coordinates of the locations of chimpanzees of the Kasekela community in Gombe National Park, Tanzania, 2000-2003</td>
</tr>
<tr>
<td>Download</td>
<td>ChimpanzeeRanges.xlsx (1.469Mb)</td>
</tr>
<tr>
<td>Download</td>
<td>README.docx (13.93Kb)</td>
</tr>
<tr>
<td>Details</td>
<td>View File Details</td>
</tr>
</tbody>
</table>

When using this data, please cite the original publication:

http://dx.doi.org/10.1098/rstb.2013.0377

Additionally, please cite the Dryad data package:


http://datadryad.org/resource/doi:10.5061/dryad.jg05d
<table>
<thead>
<tr>
<th>dc.contributor.author</th>
<th>Pusey, Anne E.</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.contributor.author</td>
<td>Pusey, Anne E.</td>
</tr>
<tr>
<td>dc.contributor.author</td>
<td>Schroepfer-Walker, Kara</td>
</tr>
<tr>
<td>dc.contributor.author</td>
<td>Schroepfer-Walker, Kara</td>
</tr>
<tr>
<td>dc.coverage.spatial</td>
<td>Africa</td>
</tr>
<tr>
<td>dc.coverage.spatial</td>
<td>Gombe National Park</td>
</tr>
<tr>
<td>dc.coverage.spatial</td>
<td>Tanzania</td>
</tr>
<tr>
<td>dc.date.accessioned</td>
<td>2013-11-05T19:47:37Z</td>
</tr>
<tr>
<td>dc.date.available</td>
<td>2013-11-05T19:47:37Z</td>
</tr>
<tr>
<td>dc.date.issued</td>
<td>2013-11-05</td>
</tr>
<tr>
<td>dc.identifier</td>
<td>doi:10.5061/dryad.jg05d/1</td>
</tr>
<tr>
<td>dc.identifier.uri</td>
<td><a href="http://hdl.handle.net/10255/dryad.52994">http://hdl.handle.net/10255/dryad.52994</a></td>
</tr>
<tr>
<td>dc.description</td>
<td>UTM coordinates of the locations of chimpanzees of the Kasekela community in Gombe National Park, Tanzania, 2000-2003</td>
</tr>
<tr>
<td>dc.relation.ispartof</td>
<td>doi:10.5061/dryad.jg05d</td>
</tr>
<tr>
<td>dc.rights.uri</td>
<td><a href="http://creativecommons.org/publicdomain/zero/1.0/">http://creativecommons.org/publicdomain/zero/1.0/</a></td>
</tr>
<tr>
<td>dc.subject</td>
<td>female transfer</td>
</tr>
<tr>
<td>dc.subject</td>
<td>dominance</td>
</tr>
<tr>
<td>dc.subject</td>
<td>infanticide</td>
</tr>
<tr>
<td>dc.subject</td>
<td>aggression to immigrants</td>
</tr>
<tr>
<td>dc.title</td>
<td>Female competition in chimpanzees</td>
</tr>
</tbody>
</table>
README

File: ChimpanzeeRanges

This file contains two sheets: AllPoints, AlonePoints

AllPoints

This sheet contains the XY UTM (zone 35M) coordinates of the locations of focal chimpanzees recorded every 15 minutes during day-long focal follows of adult chimpanzees of the Kasekela community at Gombe National Park, Tanzania in 2000-2003. It contains the following columns:

Follow Date: Date of the focal follow: Day-Month-Year

Sequence #: Consecutive 15 minute points for each follow, starting at #1 for the first observation of the day

X Coordinate: UTM coordinate, zone 35M

Y Coordinate: UTM coordinate, zone 35M

We used BIOTAS to calculate minimum convex polygons around these points.

AlonePoints

This sheet contains the XY UTM (zone 35M) coordinates of the first location on which a female was seen alone (including with her dependent offspring and adult daughters), during follows of any chimpanzee each day in 2000-2003. Females are classified as high (H), medium (M), or low (L) (see text of paper). Different females have different numbers (H1, H2, etc.). It contains the following columns:

ID: Identity of the female

Date: Date on which the female was seen alone: Day-Month-Year

X Coordinate: UTM coordinate, zone 35M

Y Coordinate: UTM coordinate, zone 35M
### All Points

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Follow Date</td>
<td>1</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>3</td>
<td>04-Jan-2000</td>
<td>2</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>4</td>
<td>04-Jan-2000</td>
<td>3</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>5</td>
<td>04-Jan-2000</td>
<td>4</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>6</td>
<td>04-Jan-2000</td>
<td>5</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>7</td>
<td>04-Jan-2000</td>
<td>6</td>
<td>791113, 9483264</td>
</tr>
<tr>
<td>8</td>
<td>04-Jan-2000</td>
<td>7</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>9</td>
<td>04-Jan-2000</td>
<td>8</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>10</td>
<td>04-Jan-2000</td>
<td>9</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>11</td>
<td>04-Jan-2000</td>
<td>10</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>12</td>
<td>04-Jan-2000</td>
<td>11</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>13</td>
<td>04-Jan-2000</td>
<td>12</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>14</td>
<td>04-Jan-2000</td>
<td>13</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>15</td>
<td>04-Jan-2000</td>
<td>14</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>16</td>
<td>04-Jan-2000</td>
<td>15</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>17</td>
<td>04-Jan-2000</td>
<td>16</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>18</td>
<td>04-Jan-2000</td>
<td>17</td>
<td>791357, 9483305</td>
</tr>
<tr>
<td>19</td>
<td>04-Jan-2000</td>
<td>1</td>
<td>791428, 9483073</td>
</tr>
<tr>
<td>20</td>
<td>04-Jan-2000</td>
<td>2</td>
<td>791428, 9483073</td>
</tr>
<tr>
<td>21</td>
<td>04-Jan-2000</td>
<td>3</td>
<td>791281, 9483032</td>
</tr>
<tr>
<td>22</td>
<td>04-Jan-2000</td>
<td>4</td>
<td>791280, 9482796</td>
</tr>
<tr>
<td>23</td>
<td>04-Jan-2000</td>
<td>5</td>
<td>791343, 9482552</td>
</tr>
<tr>
<td>24</td>
<td>04-Jan-2000</td>
<td>6</td>
<td>791343, 9482552</td>
</tr>
</tbody>
</table>

### Alone Points

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>ID</td>
<td>1</td>
<td>791428, 9483073</td>
</tr>
<tr>
<td>3</td>
<td>H1</td>
<td>2</td>
<td>791759, 9483151</td>
</tr>
<tr>
<td>4</td>
<td>H1</td>
<td>3</td>
<td>791202, 9483644</td>
</tr>
<tr>
<td>5</td>
<td>H1</td>
<td>4</td>
<td>791486, 9483168</td>
</tr>
<tr>
<td>6</td>
<td>H1</td>
<td>5</td>
<td>791830, 9483397</td>
</tr>
<tr>
<td>7</td>
<td>H1</td>
<td>6</td>
<td>790955, 9483035</td>
</tr>
<tr>
<td>8</td>
<td>H1</td>
<td>7</td>
<td>791929, 9483251</td>
</tr>
<tr>
<td>9</td>
<td>H1</td>
<td>8</td>
<td>791431, 9482531</td>
</tr>
<tr>
<td>10</td>
<td>H1</td>
<td>9</td>
<td>791482, 9483169</td>
</tr>
<tr>
<td>11</td>
<td>H1</td>
<td>10</td>
<td>791857, 9483309</td>
</tr>
<tr>
<td>12</td>
<td>H1</td>
<td>11</td>
<td>791776, 9483440</td>
</tr>
<tr>
<td>13</td>
<td>H1</td>
<td>12</td>
<td>791119, 9483710</td>
</tr>
<tr>
<td>14</td>
<td>H1</td>
<td>13</td>
<td>791380, 9483225</td>
</tr>
<tr>
<td>15</td>
<td>H1</td>
<td>14</td>
<td>791570, 9483126</td>
</tr>
<tr>
<td>16</td>
<td>H1</td>
<td>15</td>
<td>791398, 9483257</td>
</tr>
<tr>
<td>17</td>
<td>H1</td>
<td>16</td>
<td>791154, 9482274</td>
</tr>
<tr>
<td>18</td>
<td>H1</td>
<td>17</td>
<td>791540, 9483777</td>
</tr>
<tr>
<td>19</td>
<td>H1</td>
<td>18</td>
<td>792674, 9484749</td>
</tr>
<tr>
<td>20</td>
<td>H1</td>
<td>19</td>
<td>792298, 9481760</td>
</tr>
<tr>
<td>21</td>
<td>H1</td>
<td>20</td>
<td>791812, 9482150</td>
</tr>
<tr>
<td>22</td>
<td>H1</td>
<td>21</td>
<td>791459, 9482972</td>
</tr>
</tbody>
</table>
How will you share/make public your data?
YOU WANT DATA?

HERE'S SOME DATA.
Data Types to Share

• What does the grant ask for?
• **NIH - Final Research Data** - Recorded factual material commonly accepted in the scientific community as necessary to document and support research findings. (spreadsheets, images, scans of written notes if applicable, etc.)
• **OSTP** - Digitally formatted data arising from unclassified, publicly releasable research and programs.
• **NSF** – all types, documents, videos, software, etc.
Sharing Exclusions

• preliminary analyses,
• drafts of scientific papers,
• plans for future research,
• peer reviews,
• communications with colleagues,
• Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law,
• Personnel and medical information and similar information that could be used to identify a particular person in a research study.
Sharing Data

Upload to open repository; general, subject, or institutional.

- Zenodo [https://zenodo.org/](https://zenodo.org/)
- Open Science Framework [https://osf.io/](https://osf.io/)
- DataVerse [http://dataverse.org/](http://dataverse.org/)
- Search Registry of Research Data Repositories [http://www.re3data.org/](http://www.re3data.org/)
Sharing With an Article

Supplemental file with journal article or link to the upload.

– Be sure to check the contract.
– Will the data be available to the public as per OSTP if grant funded?
– Will the rights conflict with institutional ownership of the data?
– Journals often use Figshare or Dryad
Sensitive Data Access

• Researchers must request access to database, explaining research and providing IRB approval forms, e.g. registry

or

• Data must be deidentified or anonymized in some way before being made publicly available.
Concerns over bias in clinical trial reporting have stimulated calls for more open data sharing. In response, multiple pharmaceutical companies have created mechanisms for investigators to access patient-level clinical trials data. However, if and how these shared clinical trial data are being used is unknown.

**METHODS**

We evaluated how many clinical trials were publicly available to investigators through 3 open access platforms: ClinicalStudyDataRequest.com, the Yale University Open Data Access Project (YODA), and the Supporting Open Access for Researchers (SOAR) initiative. Sponsors depositing data in these platforms include GlaxoSmithKline, AstraZeneca, Boehringer Ingelheim, Eisai, Eli Lilly, Novartis, Roche, Sanofi, Takeda, Union Chimique Belge, ViV Healthcare, Johnson & Johnson, Mead Johnson, and Bristol-Myers Squibb. Company policies on what trials are shared vary and are available online, but most include all trials within certain date ranges after regulatory review and publication of results.
Sharing Sensitive Data

Reuse

What can be done with your data? Licenses can help.
ONE DOES NOT SIMPLY OWN DATA

IT DEPENDS ON POLICY
Reuse – License Your Data

• Creative Commons licenses
  https://creativecommons.org/licenses/
or use license chooser
  https://creativecommons.org/choose/

• Open Data Commons
  http://opendatacommons.org/

• Pantone Principles
  http://pantonprinciples.org/
Data Citation

• **Force 11** developed data citation format, Elsevier/Mendeley working to insure datasets have DOIs and are cited. 

• **Data Citation Index** available from VCU Libraries, and other alternative metrics, e.g. downloads in Scholars Compass, are available for data (and software).
Preserve

How long and where data will be kept.
Preserve

• How long must the data be kept?
  – Minimum 5 years after publication or final grant report.
  – Check grant and state code.

• What is the long-term value of the data?
  – If it will be in a subject repository, you can say indefinitely.
Don’t Forget Print

- Though not required for OSTP, it is needed for other agencies and backing up articles.
- Set a schedule to scan lab notebooks and other print materials (makes for a good back up and easier to share data within group).
- Print original should have similar security to digital data (i.e. good, secure storage and labelling of files).
The Journal Article Exception

Retraction Watch

“We are living in hell:” Authors retract 2nd paper due to missing raw data

with 14 comments

A 2006 paper investigating the effects of docosahexaenoic acid (DHA) and celecoxib on prostate cancer cells has been retracted because it appears to contain panels that were duplicated, and the authors could not provide the raw data to show otherwise.

This is the second paper the authors have lost because they couldn’t furnish the original data to defend their work against allegations of image manipulation. The reason: the Institute for Cancer Prevention in New York, where the authors did the work, shut its doors abruptly in 2004, co-author Bhagavathi A. Narayanan told us. (The institute closed thanks to $5.7 million in grant that was misspent, the New York Post reported at the time.)

Recently, some of Narayanan’s papers have been questioned on PubPeer: her work has been the subject of an investigation at New York University, where Narayanan is now based.

Narayanan told us that the criticism of their work has deeply affected her and her co-authors:

“We are living in hell.”
Narayanan said that at the time the work was done, over a decade ago,

There were no rules that you had to keep the data. There was no Retraction Watch.

When we asked Narayanan about her work, she told us “of course science papers have mistakes.” In regards to the comments on PubPeer, she added:

It’s discrimination, [it’s] jealousy, it is targeting somebody. Most of the PubPeer comments were meritless. They just want to hurt the people... This is not a pleasant experience to share... This is, at the expense of someone’s dead body, eating the other person’s flesh.
Stored Example

Final MS for deposit

Data to support figure and images
Research doesn't leave me any time for exercise!

Do you know a way I can get a good cardio workout?

All your data was lost when the computer crashed.

AAAAAAAAAAAA
AAAAAAAAAAAA
AAAAAAAAAAAA

There you go.