Public Access to Articles and Data for Funding Agencies

Aaron McCollough and Heidi Imker
OSTP Public Access Memo

INCREASING ACCESS TO THE RESULTS OF FEDERALLY FUNDED SCIENTIFIC RESEARCH

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.
OSTP Public Access Memo

• maximize the impact and accountability of the Federal research investment
• accelerate scientific breakthroughs and innovation, promote entrepreneurship, and enhance economic growth and job creation.

Each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government.

http://go.illinois.edu/OSTP_Memo
Overview of Implementation Plans

Retrieved 17:18, Jun 03, 2015

<table>
<thead>
<tr>
<th>Funder</th>
<th>Published Outputs</th>
<th>Data</th>
<th>Time Limits</th>
<th>DMP</th>
<th>Access/Sharing</th>
<th>Long-term Curation</th>
<th>Monitoring</th>
<th>Guidance</th>
<th>Article Repository</th>
<th>Data Repository</th>
<th>Costs</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASPR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDC</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DOT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NASA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NIST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NOAA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USDA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USAID</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>USGS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Last updated: May 7, 2015
Crowd-sourced Google Doc

Comprehensive resource for article and data requirements expected from the implementation plans.

Spreadsheet

Complementary FAQ

Articles
To the extent feasible and consistent with law; agency mission; resource constraints; U.S. national, homeland, and economic security; and the objectives listed below, the results of unclassified research that are published in peer-reviewed publications directly arising from Federal funding should be stored for long-term preservation and publicly accessible to search, retrieve, and analyze in ways that maximize the impact and accountability of the Federal research investment.
Publication Deposit Mandate

• Requires making federally funded research openly available.
• Requires deposit whether or not the original publication is open access.
• Is fulfilled in most cases through deposit to a central government repository (like PubMed), but may for some agencies be fulfilled by open access publication on the publisher’s website, or deposit into an institutional repository.
• Generally will need to be done within a year of publication or of acceptance for publication.
Sample Implementation Plan

National Science Foundation (NSF)

- Authors must submit either the version of record or the final accepted peer-reviewed manuscript to the [DOE’s Public Access Gateway for Energy and Science (PAGES) repository](https://doi.org/10.26134/123456) in PDF/A format that should be available for download, reading, and analysis free of charge no later than 12 months after initial publication, with machine-readable metadata available at initial publication. Also: *Authors will regularly report on the status of their deposit of materials using quarterly reports, where applicable; annual progress reports; and final progress reports.*
Effective Dates for Publication Deposit

Unspecified: DOT, DOD, USAID, USGS
These dates generally indicate when the agency plans to incorporate requirements into the application process, but several indicate the policy will apply to existing grants that are still in progress, so some dates may be earlier in practice.
What Do You Deposit

- Any peer-reviewed research publication.
- Generally either the final accepted manuscript or, when allowed by the publishing contract, the final published version.
## Where is the Publication Deposited?

<table>
<thead>
<tr>
<th>Agency</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>AHRQ, ASPR, FDA, NIH, NIST, VA</td>
<td>PubMed</td>
</tr>
<tr>
<td>USDA</td>
<td>PubAg</td>
</tr>
<tr>
<td>CDC</td>
<td>CDC Stacks</td>
</tr>
<tr>
<td>DOD</td>
<td>Defense Technical Information Center</td>
</tr>
<tr>
<td>DOE</td>
<td>1) Article publisher, 2) local repository [e.g., IDEALS at U Illinois], or 3) OSTI [note: metadata and a DOI for the version of record must be provided to OSTI for all publications for indexing purposes, and will be searchable through PAGES, and OSTI must receive copies of all articles to keep in a dark archive for use if option 1 or 2 fails]</td>
</tr>
<tr>
<td>NASA</td>
<td>NASA-branded version of PubMed</td>
</tr>
<tr>
<td>NOAA</td>
<td>New NOAA Institutional Repository</td>
</tr>
</tbody>
</table>
Who Deposits the Publication?

• Many journal publishers will deposit either the final accepted version of the article or the final published version in PubMed for NIH-funded research, but not all.
• Details, including participating journal and publisher lists: [https://publicaccess.nih.gov/submit_process.htm](https://publicaccess.nih.gov/submit_process.htm)
• Those that do may not all extend this service to research funded by newer agency mandates, either articles going into PubMed or other agency options (PubAg, CDC Stacks, etc).
When Do You Deposit?

• For most agencies, deposit of articles is required **within 12 months of publication**, although many also specify deposit (or at least initiation of the process) upon acceptance for publication.

• Be sure to check your journal’s **author agreement** before you sign to ensure that deposit of your work will not conflict.

• Some agencies explicitly indicating that compliance will be tracked and affect eligibility for future awards.
Making it Work

- Tiered compliance obligations (individual researcher & University)
- The New U of I Faculty **Open Access Policy** (no waiver permitted for covered federal grant publications)
- The University of California’s Example and Experience
- Enterprise-level efforts at scale solutions (CHORUS, SHARE, & PURE)
Scholarly Commons

• Learn more about your specific funder(s) policies: http://tinyurl.com/p2p4es7

• Help you evaluate your publisher agreement to make sure there are no conflicts with deposit requirements.

• May be able to help you determine if your publisher will automatically deposit your work for you.

• Work with subject liaison to determine high impact OA publishing options.

e-mail sc@library.illinois.edu
Research Data
To the extent feasible and consistent with law; agency mission; resource constraints; U.S. national, homeland, and economic security; and the objectives listed below, the results of unclassified research that are published in peer-reviewed publications directly arising from Federal funding should be stored for long-term preservation and publicly accessible to search, retrieve, and analyze in ways that maximize the impact and accountability of the Federal research investment.
To the extent feasible and consistent with law; 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 for long-term preservation and publicly accessible to search, retrieve, and analyze in ways that maximize the impact and accountability of the Federal research investment.

2 https://www.whitehouse.gov/omb/circulars_a130_a130trans4/
Long-Term Preservation

DOD Implementation Plan
... assessment on the inherent value of the data and provide a recommendation on whether to transfer the data elsewhere or to abandon the notion of preservation. Pg 15

NIH Implementation Plan
The costs and benefits of data management should be considered in policy development and implementation. It is important to note that not all digital scientific data need to be shared and preserved. A fundamental determinant of the need to share and preserve data is the further insight to be gained from increased public access. Pg 24

NSF Implementation Plan
... rarely does NSF expect that retention of all data that are streamed from an instrument or created in the course of an experiment or survey will be required. Pg 15
OSTP Key Points for Research Data

- free public access to research data
- data management planning
- citation of datasets
- private sector collaborations
- education and workforce development
- assessment of data preservation needs
- development of sustainable solutions
“requiring researchers to better account for and manage the digital data resulting from federally funded scientific research”

• Data management plans will be come compulsory
• Providing public access to data will become more routine
What’s in Common

• **DMP coverage**
  - What data is produced, how will the data will be accessible, how the data will be preserved, what access restrictions are expected

• **DMPs currently required by NSF, DOE, and USGS**

• **DMP impending requirements**
  - 8 in Oct 2015 (AHRQ, ASPR, CDC, DOD, DOT, FDA, NASA, NIH,* NIST)
  - 2 in Jan 2016 (USDA, NOAA)

* NIH moving from a single paragraph “Data Sharing Plan” required only from grants with >$500K direct costs per year to DMP for all research applications
Specific language can vary between agencies, between sub-agencies, and even between different funding announcements.
What’s Very Variable
### Promoting an open research

#### Data transparency

<table>
<thead>
<tr>
<th>Data transparency</th>
<th>Citation standards</th>
<th>Research materials transparency</th>
<th>Design and analysis transparency</th>
<th>Preregistration of studies</th>
<th>Preregistration of analysis plans</th>
<th>Replication</th>
</tr>
</thead>
<tbody>
<tr>
<td>Article states whether data are available and, if so, where to access them.</td>
<td>Journal describes citation of data in guidelines to authors.</td>
<td>Article states whether materials are available and, if so, where to access them.</td>
<td>Journal requires adherence to design transparency standards for review and publication.</td>
<td>Journal encourages preregistration of studies and provides link in article to preregistration if it exists.</td>
<td>Journal encourages preregistration plans and provides link in article and certification of meeting preregistration badge requirements.</td>
<td>Journal encourages submission of replication studies.</td>
</tr>
<tr>
<td>Data must be posted to a trusted repository. Exceptions must be identified at article submission.</td>
<td>Article provides appropriate citation for data and materials used, consistent with journal's guidelines.</td>
<td>Materials must be posted to a trusted repository. Exceptions must be identified at article submission.</td>
<td>Journal requires and enforces adherence to design transparency standards for review and publication.</td>
<td>Journal encourages preregistration of studies and provides link in article to certification of meeting preregistration badge requirements.</td>
<td>Journal requires preregistration of studies with analysis plans and provides link and badge in article to meeting requirements.</td>
<td>Journal encourages submission of replication studies and conducts blind review of results.</td>
</tr>
<tr>
<td>Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.</td>
<td>Article is not published until appropriate citation for data and materials is provided that meets journal's guidelines.</td>
<td>Materials must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.</td>
<td>Journal requires preregistration and provides link and badge in article to preregistration before publication.</td>
<td>Journal requires preregistration of studies and provides link and badge in article to preregistration before publication.</td>
<td>Journal uses Registered Reports as a submission option for replication studies with peer review before observing the study outcomes.</td>
<td></td>
</tr>
</tbody>
</table>

---

Journal Requirements for Data

*e.g. Science* General Information for Authors
(*but PLOS, Evolution, others...*)

**Data and materials availability** All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of *Science*.

**First preference:** Their website (but 25 MB limit) or existing established repository

**Second preference:** “…authors will commonly be required to sign an agreement to make a copy of the data[base], as it exists at the time of publication, publicly available on their own Web sites for a period of at least five years after the paper's publication date.”

http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#unpublished
The Irish Potato Famine Pathogen Phytophthora infestans Translocates the CRN8 Kinase into Host Plant Cells
van Damme et al
PLoS Pathogens 2012

Posted on PubPeer
1. New expectations for greater transparency with associated technologies, platforms, and policies for enabling it.
2. Fraction targeted likely will be small, but be prepared for...
Effective Dates?
Data Management Plans Only!

AHRQ, ASPR, CDC, DOD, DOE, FDA, NASA, NIST, VA
Oct ‘15

DOE
Oct ‘14

USGS
Feb ‘15

NIH
Dec ‘15

NSF, NOAA, USAID
Jan ‘16

Unspecified: DOT, USDA

2013 | 2014 | 2015 | 2016

NSF
2011

Today
What Data is Deposited?

Data is defined, consistent with OMB circular A-110, 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 ...

- lab notebooks
- preliminary analyses
- drafts of scientific papers
- plans for future research
- peer review reports
- communications with colleagues
- physical objects, such as lab specimens
Where is the Data Deposited?

• Support established, disciplinary repositories
Who Deposits the Data?

Researcher
When is the Data Deposited?

• most agencies say with article publication
The Research Data Service (RDS) provides the Illinois research community with the expertise, tools, and infrastructure necessary to manage and steward research data.
Research Data Service

Expertise

• Knowledge around data policies, tools, resources, archiving, and preservation
• Consultation and workshops for data management planning and implementation

Tools

• Data Management Plan creation wizard (DMPTool.org)
• Tools for data citation (DOI minting)

Infrastructure (in progress)

• Centralized, managed storage for active (“working”) data
• Solutions for public access to research data
DMP Consultation – Option 1

https://dmptool.org/

- online wizard for creating a ready-to-use DMPs
- help text, links to resources, and Illinois-specific suggestions
- login with Illinois NetID and password to automatically create an account
- create, save, and edit plans
DMP Consultation – Option 2

Email to researchdata@library.illinois.edu

We’ll request:
- link to the application instructions
- proposal abstract
  - confidential – simply to provide direction to most appropriate resources
- deadline that the application needs to be OSPRA

We provide:
- review in 1-2 business days (usually 1 and frequently outside of work hours)
  - more time = more comprehensive review
  - guidance including suggested resources, revisions, or language

We don’t provide:
- a DMP written wholesale
- funding guarantees
Contact

Aaron McCollough
amccollo@illinois.edu

Scholarly Commons
http://www.library.illinois.edu/sc/

Heidi Imker
imker@illinois.edu

Research Data Service
http://researchdataservice.illinois.edu/

... and our thanks to LIS Librarian Dan Tracy for slides!