

## SPARC Resource on Federal Data Sharing Requirements and Plans

Jonathan Petters

jon\_petters

JHU Data Management Services

Email: [datamanagement@jhu.edu](mailto:datamanagement@jhu.edu)

Website: <http://dmp.data.jhu.edu/>

# Credit

## Johns Hopkins

- Amanda Brite
- David Fearon
- Barbara Pralle
- Winston Tabb

## SPARC

- Heather Joseph
- Nick Shockey
- Greg Tananbaum
- Josh Currie (H&H  
Web Development)

# Outline

- Goal, Context and Scope
- Methodology
- Web Resource Functionality
- High-level Preliminary analysis (Observations)
- Closing Comments
- Q&A

# Goal

- Organize and surface relevant expectations and anticipated requirements found in the Federal Agency Public Access Plans as well as current and future policies related to public access of digital research data products

# Context



THE DIRECTOR

EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF MANAGEMENT AND BUDGET  
WASHINGTON, D.C. 20503

May 9, 2013

MEMORANDUM FOR THE DIRECTOR

M-13-13

FROM: John P. Holdren  
Director

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Increasing Access to Information

FROM: Sylvia M. Burwell  
Director

Steven VanRoekel  
Federal Chief Information Officer

Todd Park  
U.S. Chief Technology Officer

Dominic J. Mancini  
Acting Administrator, Office of Information and Regulatory Affairs

SUBJECT: Open Data Policy—Managing Information as an Asset

# Goal and Scope

- Organize and surface relevant expectations and anticipated requirements found in the Federal Agency Public Access Plans as well as current and future policies related to public access of digital research data products
  - Publication requirements out of scope
  - Private foundations out of scope
  - Focused on extramural research programs

# Agencies

## Data Management Services



User of agency logos is in no way intended to imply endorsement by any agency of the SPARC web resource or this presentation.

# datasharing.sparcopen.org

RC\*

Who We Are

What We Do

Why It Matters

Become a Member



## Browse Data Sharing Requirements by Federal Agency

This community resource for tracking, comparing, and understanding both current and future U.S. federal funder research data sharing policies is a joint project of SPARC & Johns Hopkins University Libraries. Click the icons below to select up to three agencies to view or compare. Click [here to download](#) the full data set.



Search for an agency...



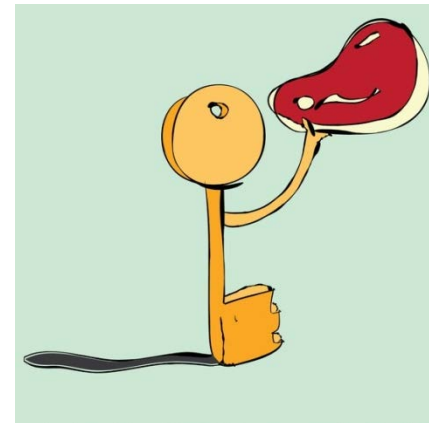


# For Each Agency

- Investigated Public Access Plan and/or existing data sharing and management policies
- Distinguished between current policy ('Currently') and plans ('Looking Forward')
- Consistent set of facets by which we examined relevant data sharing practices and policies

# Facets

- Thirteen facets chosen, based on our experience, discussions with SPARC
- For each facet, pinpointed passages that, in 100-ish words, communicated most pertinent information to researchers, research libraries and other key stakeholders



# For Each Agency

- Extracted passages from policies and plans only (not supplementary materials like FAQs)
- Sometimes required light re-ordering or paraphrasing, but intent was to NOT editorialize or interpret passages
- Included links to relevant supplementary documents or resources

# Example – CDC Content



## Exclusions and limitations in Policy Approach

CURRENTLY | LOOKING FORWARD

The CDC Data Policy has the following within its scope: "[d]ata collected by CDC using federal resources, [d]ata collected for CDC by other agencies or organizations..., [and d]ata reported to CDC (e.g., by a state health department)". Data are defined as "scientific records which are as accurate and complete as possible". "This policy does not cover data shared with CDC but owned by other organizations (e.g., data provided to CDC by a managed care organizations, preferred provider organizations, or technology firms for a specific research project)" (43).

## When Data are to be Publicly Shared

CURRENTLY | LOOKING FORWARD

Data "underlying the conclusions presented in peer-reviewed research papers freely available in machine-readable formats at the time of initial publication" unless the data has already been made available to the public or they are excluded (13). "In cases where a minimal dataset is released coincident with the research paper, but CDC intends to release a more detailed dataset after it is cleaned, documented, and vetted, the initial dataset release can be followed with a more complete data release. The second release will take place according to CDC's standard timeline for release of

# Compare Agencies



## CURRENTLY | LOOKING FORWARD

"CDC believes that public health and scientific advancement are best served when data are released to, or shared with, other public health agencies, academic researchers, and appropriate private researchers in an open, timely, and appropriate way" (42).

"The CDC Data Policy is guided by the following principles: accountability, privacy and confidentiality, stewardship, scientific practice, efficiency, and equity" (19).

The CDC/ATSDR Policy on Releasing and Sharing Data (hereafter "CDC Data Policy") details several benefits of releasing or sharing CDC data (43-44).

"Public access to the results of FDA-funded scientific research furthers the agency's public health mission" (1). "Facilitating the free flow of information underlying the agency's decision-making and advances in regulatory science, to the extent permitted by law, allows the public, Congress, media, and industry to better understand FDA's decisions and the scientific basis for its regulatory decision-making" (1). "While the agency has embraced the values of openness and transparency, the agency's role as a regulator does present restrictions on the full and free disclosure of scientific data" (1-2).

Implementing OSTP public access requirements is consistent with NSF's 2014-2018 Strategic Plan to "[t]ransform the frontiers of science and engineering", "[s]timulate innovation and address societal needs through research and education", and "[e]xcel as a federal science agency" (2-3).

## Principles

# References and Dates



## Documents Cited

CDC Plan for Increasing Access to Scientific  
Publications and Digital Scientific Data  
Generated with CDC Funding

CDC/ATSDR Policy on Releasing and Sharing  
Data

Plan to Increase Access to Results of  
FDA-Funded Scientific Research

SMG 2126.4 FDA Staff Manual Guides, Volume  
III - General Administration; External Relations;  
Access to Results of FDA-Funded Scientific  
Research

NSF's Public Access Plan: Today's Data,  
Tomorrow's Discoveries

## Content Creator and Date Completed (ISO 8601)

Johns Hopkins Data Management Services  
2015-11-23

Johns Hopkins Data Management Services  
2016-01-24

Johns Hopkins Data Management Services  
2015-09-14

# PRELIMINARY Analysis

(Or general overall impressions)

- All agencies working with the same directives - OSTP 2013 and Open Data Policy (M-13-13)
- However, each agency has a different
  - Mission
  - funded community
  - Resources invested in existing infrastructure (including with respect to data management and sharing)

# Principles



## CURRENTLY | LOOKING FORWARD

"Public access to the results of FDA-funded scientific research furthers the agency's public health mission" (1). "Facilitating the free flow of information underlying the agency's decision-making and advances in regulatory science, to the extent permitted by law, allows the public, Congress, media, and industry to better understand FDA's decisions and the scientific basis for its regulatory decision-making" (1). "While the agency has embraced the values of openness and transparency, the agency's role as a regulator does present restrictions on the full and free disclosure of scientific data" (1-2).

How does public access fit into agency mission?



# Implementation Approaches



CURRENTLY | LOOKING FORWARD

Intramural reqs first

The FDA Staff Manual Guide on Access to Results of FDA-Funded Scientific Research (hereafter the "Guide") enacts data access and management requirements for FDA-funded intramural research initiated or renewed after December 29, 2015, and FDA-funded extramural research initiated or renewed after October 1, 2016 (Guide 3). "Given the presumption of openness of agency data, the

-----

datasets to allow disclosure" (Guide 9). "The Office of Public Health Strategy and Analysis...will collaborate with the Office of Scientific Integrity to implement this Guide" (12). "The Senior Science Council...will review and provide expert input" (Guide 12-13).

Responsible parties/process noted, varies

# Exclusions and Limitations in Policy Scope



CURRENTLY | LOOKING

Research data definition  
(varies, sometimes  
includes software)

Research data are defined as per the OMB circular A-110, covering "digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings" (4).

Research data may be reviewed for public release, through an examination "for compliance with established national and DoD policies and to determine that it contains no classified or export-controlled information" (6). "For the purposes of this draft proposed plan, data will not be publicly releasable if release would compromise the ability to file for intellectual property protection on any invention arising from the data" (3). "Data that is not approved for public release will not be included under this plan" (15). Further, data containing personally identifiable information is restricted (5).

The omission of data for the any qualifying reason must be addressed in the DMP (4).

Variation in exclusions

# When Data Are to Be Shared



Data generally to be shared upon publication

CURRENTLY | LOOKING FORWARD

"AHRQ expects the timely release and sharing of data to be no later than the acceptance for publication of the main findings from the final dataset. The specific time will be influenced by the nature of the data collected" (Section 3). "If data from large epidemiologic or longitudinal studies are collected over several discrete time periods or waves, it is reasonable to expect that the data would be released in waves as data become available or main findings from waves of the data are published. Researchers will be directed to work with AHRQ staff and its contracted commercial repository to deposit data upon or prior to the embargo period" (Section 3).

How much? 'Underlying publication'

Embargo of data release sometimes an option

# How Data Are to be Shared



Approaches to develop data sharing infrastructure vary

unless an exemption is granted by the NOAA Program" (NOAA Directive 10). "Approved submission of data to [the NOAA National Centers for Environmental Information], or to a publicly-accessible data repository approved by the funding program, shall be sufficient to satisfy the data accessibility requirement" (NOAA Directive 4). "[H]owever, NCEI is not obligated to accept all submissions and may charge a fee, particularly for large or unusual datasets."

Use existing repositories, create new ones, create data catalogs

What is a repository? What makes it appropriate for deposit?

# Metadata/Documentation to be Included



CURRENTLY | LOOKING

Plans/policies note existing metadata standards

DoD provides a definition for metadata, citing Dublin Core Metadata Initiative (DCMI) and Department of Defense Discovery Metadata Standard (DDMS) as two examples of established standards (5).

CURRENTLY | LOOKING FORWARD

"DoD will develop requirements for the submission of metadata to DTIC. The metadata for scientific data will include, at a minimum, the common core metadata schema in use by the federal government, found at <https://project-open-data.cio.gov/>" (12).

Plans to use Project Open Data common core

# Data Citation and Attribution



CURRENTLY | LOOKING

Data citation standards  
emerging and evolving

"Currently, there is no NIH-wide system for providing consistent identification and attribution for NIH-funded data that are made publicly available. While most NIH-supported data repositories provide unique identifiers for submitted data sets, practices for attribution and citation vary from one system to another." (39)

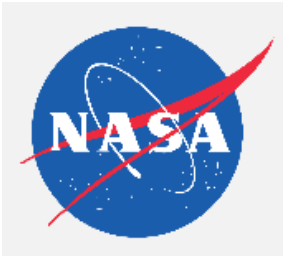
CURRENTLY | LOOKING

Agencies encourage (a few  
require) data citation usage

"NIH will explore ways to advance data as a legitimate form of scholarship through data citation and other means", including the creation of an identifier analogous with PMCID to link data with biomedical literature via PubMed records to facilitate full citations (40).



# Data Licensing



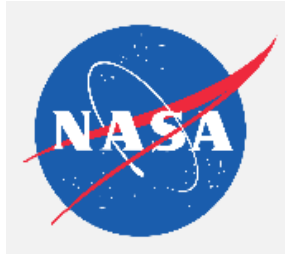
Few agencies give data  
licensing guidance currently

NASA will expand upon its existing data policies regarding "policies and provisions for reuse, redistribution, and the production of derivatives" (9).

Some (but not all) will explore approaches  
in guiding researchers in this area

# Data Management Planning

Data Management Services



Data management plan to

CURRENTLY | LOOKING FORWARD

All research proposers require a DMP with the proposal or project plan, describing "how the proposed research plan conforms to NASA policy on the dissemination and sharing of research results". The DMP will address:

- types of data to be produced
- the standards to be used for data and metadata format and content
- policies for accessing and sharing the data, including provisions for restricting material
- policies and provisions for re-use, re-distribution, and the production of derivatives
- plans for providing access to the data used in any science publication
- plans for archiving and preserving data, as

Conceptual requirements are found on (6).

Storage during project included?



# Data Preservation



Will assess needs for preservation,  
consider cost of doing so

CURRENTLY | LOOKING FORWARD

"NIST will assess the long-term needs for preservation of scientific data in fields that the agency supports, and outline options for developing and sustaining repositories for scientific data in digital formats, taking into account the efforts of public- and private-sector entities." (11-12)

How long to sustain data access?  
What does sustainability of data  
access mean?

# Partnerships in Policy Enactment



Have participated in interagency meetings, public consultant sessions

CURRENTLY | LOOKING FORWARD

"For elements of the Department for which the collection of research data is not already practiced, DOE sponsoring offices should consult with their research communities through public forums such as Federal Advisory Committee Meetings and public announcements to identify which research data are appropriate for the DOE to collect or otherwise include in the public listing of agency data, required by the Open Data Policy, and suitable mechanisms for doing so" ([Roles and Responsibilities: DOE Sponsoring Research Offices](#)).

Many agencies have existing channels to garner community input

# Compliance for Policy Enforcement



Agencies require DMP at proposal stage

CURRENTLY | LOOKING FORWARD

[Fa] "DOT will...develop policies to ensure new awards to researchers or institutions are not finalized until the awardee has successfully satisfied all terms of previous awards from DOT, including compliance with publication requirements and Data Management Plans" (7). "Awardees and Their Institutions" will "[e]nsure all terms and conditions of awards are met, including the submission of Research Project Records...and Digital Data Sets that arise in whole or in part with funding received through a DOT-managed contract,

Compliance checking approaches vary  
(proposal systems already exist)

# Guidance and Support for Researchers



Data management and sharing costs can be included in proposals

CURRENTLY | LOOKING FORWARD

"Guidance for the inclusion of appropriate costs for data management and access in extramural funding proposals" (17). Instructions and FAQs will be developed for proposal submission via grants.gov (17). Application templates may be developed, and project data management plan evaluation rubrics may be developed "to ensure proper assessment of plan merits" (17).

FAQs and DMP templates developed or forthcoming for many agencies

# When Using Multiple Funding Sources



Some agencies have statements similar to NIST

CURRENTLY | LOOKING FORWARD

"For activities funded by multiple sources with differing public access requirements, the provisions of this plan will apply unless otherwise specified by NIST in its funding documents" (2).

Others do not address this facet

# PRELIMINARY Analysis

(Or general overall impressions)

- All agencies working with the same directives - OSTP 2013 and Open Data Policy (M-13-13)
- However, each agency has a different
  - Mission
  - funded community
  - Resources invested in existing infrastructure (including with respect to data management and sharing)

# Closing

- Use the resource! Share it!
- SPARC exploring alternatives in the management of the content as new data sharing policies emerge
- Contact SPARC with questions and feedback at [sparc@sparcopen.org](mailto:sparc@sparcopen.org) or via the website

**Thanks for Listening!**  
**Questions?**

Jonathan Petters

jon\_petters

**JHU Data Management Services**

Email: [datamanagement@jhu.edu](mailto:datamanagement@jhu.edu)

Website: <http://dmp.data.jhu.edu/>



# Other Relevant Resources

- [OSTP](#) and [CENDI](#) – list of Public Access Plans
- [Crowdsourced OSTP spreadsheet](#)
- Research data service units in academic libraries and other institutions

# DOE Principles

Data Management Services



## Principles and Implementation Approach

### CURRENTLY | LOOKING FORWARD

"The Department affirms that the following principles for the management of digital research data support its mission [[see Unlabeled Introduction](#)]:

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
- Sharing and preserving data are central to protecting the integrity of science...data sharing should make digital research data available to and useful for the scientific community, industry, and the public.
- Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning." ([Principles](#)).

# NSF – How Data Are to be Shared



## How Data are to be Publicly Shared

CURRENTLY | LOOKING FORWARD

Data underlying findings reported in a journal article or conference paper should be made available in accordance with publisher policies and what was laid out in the proposal DMP (6). PIs are expected to share their data, both digital and physical, "at no more than incremental cost and within a reasonable time" (7).

CURRENTLY | LOOKING FORWARD

In the future, the NSF plans to explore whether all data should be made available at the time of article publication (6) NSF will "modify internal systems and the NSF public web page to support search of data sets" (18) and "will consult with the research communities to develop discipline-specific guidance and best practices" (12)

# Facets

- Principles and Implementation Approach
- Exclusions and Limitations in Policy Scope
- When Data are to be Publicly Shared
- How Data are to be Publicly Shared
- Metadata and/or Documentation to be Included with Data
- Data Citation and Attribution (subcomponent of metadata/documentation)
- Recommendations or Requirements for Terms of Data 'Licensing' (e.g. Re-use/Re-distribution/Derivatives)

# Facets

- Data Management Planning
- Data Preservation
- Current or Planned Partnerships in Policy Enactment
- Compliance Measures for Policy Enforcement
- Guidance and Support to Researchers in Following Policy
- Guidance on Data Sharing/Management for Researchers using Multiple Funding Sources