COGR
Council On Governmental Relations

Public Access Panel

Panel will be moderated by Jackie Bendall, COGR

Panel members

Jim Luther, AVP Finance & Compliance Officer, Duke University
Stephanie Endy, AVP for Research, Case Western University
Tobin Smith, VP for Policy, Association of American Universities (AAU)

Agenda

- Regulatory Overview of Public Access Requirements
  - 2013 OSTP Requirement
  - Current Status of various agencies
- COGR & University Perspective
  - NIH RFI: Strategies for NIH Data Management, Sharing, and Citation
- AAU Public Access Working Group
- Discussion
  - Implementation and Costing implications

The OSTP Public Access Memo and its Repercussions

Stephanie Endy

Case Western Reserve University

Increasing Access to the Results of Federally Funded Scientific Research

- February 2013
- Addressed to Heads of Executive Departments and Agencies
- Directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan

Elements of the Required Plans

Plan must address both scientific publications and digital scientific data and:

a) Leverage existing archives
b) Strategy for improving public's ability to locate and access data
c) Approach for optimizing search, archival, and dissemination features encouraging access and operability and long-term stewardship of results
d) Plan for notifying awardees and federal researchers of their obligations
e) Agency strategy for measuring and enforcing compliance
f) Resources within existing budget to implement
g) Timeline for implementation
h) Identification of special circumstances preventing agency from meeting these objectives

Data

“For purposes of this memorandum, 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 laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens.”

Each plan shall:

- Maximize access (protecting confidentiality, recognizing proprietary interest and IP considerations, and preserving balance of long-term preservation vs. cost)
- Make sure researchers develop data management plans addressing long-term preservation and access in digital formats
- Allow for costs for data management and access in proposals
- Ensure appropriate evaluations of data management plans
- Include mechanisms for compliance
- Promote use of public databases
- Encourage cooperation with private sector
- Develop approaches for appropriate attribution of data sets
- Support training and education related to data management, analysis, storage, preservation, and stewardship
- Provide for assessment of long-term preservation needs including options for developing and sustaining repositories.

SPARC

Sparcopen.org, “Who We Are”

“SPARC (the Scholarly Publishing and Academic Resources Coalition) works to enable the open sharing of research outputs and educational materials in order to democratize access to knowledge, accelerate discovery, and increase the return on our investment in research and education.”

Agencies Without Plans

- Department of Commerce
- HHS/Assistant Secretary for Preparedness and Response*
- HHS/NIH*
- Department of Housing and Urban Development
- Department of the Interior
- Department of Labor
- Institute for Museum and Library Services

Agency	Publications
Department of Agriculture	USDA specific repository Requires submission of final peer-reviewed manuscripts
Department of Defense	DoD specific repository Final peer-reviewed manuscripts
Department of Education	ERIC, Institute of Education Sciences repository Peer-reviewed publications in form of final manuscript
Department of Energy	PAGES, DoE specific repository Final peer-reviewed manuscripts or articles
Health and Human Services	PubMed, but also formerly Stacks for CDC Attempting to unify all four divisions affected (NIH, CDC, FDA, and AHRQ)
Homeland Security	Use PubMed Central
Department of Transportation	DOT specific repository (National Transportation Library) Final peer-reviewed publications Scientific Research Project written deliverables
Veteran's Affairs	Plans to use PubMed Central; considers impact of requirement primarily on their own hospitals rather than extramural
Environmental Protection Agency	Use of PubMed Central Concerned about publication copyright in particular
NASA	NASA PubSpace (managed by PubMed Central) Copies of peer-reviewed scientific publications and associated data – excluding publications with restrictions
NIST	Final peer-reviewed manuscripats NIST specific repository
NOAA	Publications include technical reports and professional papers issued or sponsored by NOAA Otherwise final pre-publication manuscripts Repository to be identified by NOAA
NSF	Final peer-reviewed manuscripts NSF-PAR (Public Access Repository) joint with DoEnergy (note conflict in plans with DoE published plan)
Smithsonian	Final peer-reviewed manuscript Smithsonian managed or approved repository (Smithsonian Research Online or Clearinghouse for Open Research of the United States)

M-13-13

- May 2013
- Memo to agencies
- Open Data Policy – Managing Information as an Asset
- “Specifically, this Memorandum requires agencies to collect or create information in a way that supports downstream information processing and dissemination activities.”
- Includes “open data” presumption and accessibility

Agency	Data
Department of Agriculture	We will write a policy addressing all the requirements
Department of Defense	Data Management Plan required; There will be metadata requirements; Data can be stored centrally or institutionally
Department of Education	Data Management Plan required Contractors must provide data sets
Department of Energy	Data Management Plan required
Health and Human Services	Data Sharing Plans required; Data Management Plans to be required in future; Each sub-agency is slightly different
Homeland Security	Data Management Plan required
Department of Transportation	Data Management Plan required; Plan must include deposit of data set in a repository
Veteran's Affairs	Data Management Plan required
Environmental Protection Agency	Data Management Plan required
NASA	Data Management Plan required NASA specific repository
NIST	Data Management Plan required Building a repository with index
NOAA	Their own definition of data Data Sharing Plan required Repository to leverage existing data centers; data archiving according to their own priorities
NSF	Data Management Plan required
Smithsonian	Data associated with a publication swept under publications Public repository with planned long term migration as needed to support the community

COGR and University Perspective

Jim Luther

Duke University

COGR & University Perspective

- General Focus
  - Implementation, Operationalization & Burden
  - Costing: Who Pays?

- NIH RFI: Strategies for NIH Data Management, Sharing, and Citation Notice Number: NOT-OD-17-015
  - COGR Response
  - Institutional responses - Sample

NIH RFI: Strategies for NIH Data Management, Sharing, and Citation; Notice Number: NOT-OD-17-015

- Released: 11/14/16 // Response Due: 1/19/17
- Seeks public comments…(1) how digital scientific data… should be managed, and to the fullest extent possible, made publicly available; and, (2) how to set standards for citing shared data and software (RPPR & Applications).

- Further, effective data sharing relies upon appropriate identification, adoption, and crediting of good data management and sharing practices… NIH is adopting **"FAIR" (Findable, Accessible, Interoperable, and Reusable);** http://www.nature.com/articles/sdata201618).

NIH RFI - Reference

http://www.nature.com/articles/sdata201618#abstract

- Specific and Existing Repositories
  - Genbank, Worldwide Protein Data Bank (wwPDB), and UniProt in the life sciences; Space Physics Data Facility (http://spdf.gsfc.nasa.gov/) ,etc.

- General-purpose data repositories
  - Dataverse, FigShare, Dryad, Mendeley Data , Zenodo, DataHub, DANS, EUDat.
    - …wide variety of formats, generally do not attempt to integrate or harmonize the deposited data, and place few restrictions on the descriptors of the data deposition…
    - Resulting data ecosystem, therefore, appears to be moving away from centralization….. thereby exacerbating the discovery and re-usability problem for both human and computational stakeholders.

RFI Structure

- How digital scientific data… should be managed
  - Highest Priority Type of Data and Value
  - Length of Time
  - Barriers
  - Other

- How to set standards for citing shared data and software (RPPR & Applications).

AAU Public Access Working Group

Tobin Smith
Association of American Universities

Discussion

Implementation and Costing Implications

Jim Luther

Duke University

Implementation & Operationalization

Flexibility v. Consistency

- Unknown/unclear requirements at proposal
- Sponsor harmonization (format, metadata, duration, etc.)
- Faculty Burden: DMP development, policy compliance, ongoing support after data is accessible
- Integration into DMP (too much or too little…)
- Institution's IT strategy and ability to expand and contract based on life of award (short-term) and long-term repository needs

Issue and Discussion
Costing Implications

- Costing Mechanisms
  - Externally Provided Solution
    - Federal, Publisher, Association, etc.
  - Institutionally Provided Solution
    - Capped rate vs. Uncapped rate
  - Direct Charge to federal sponsor
  - Charge to secondary user

Institutional Example

- Light microscopy: 30-100Gb/experiment, 100 experiments/researcher, 20-30 researcher/yr. Projection: 300Tb/yr
- CryoEm: Potential storage needs of ~400Tb/yr

Size / Timeframe	Annually	5 Years (one time)	7 years (one time)	Perpetual (one time)
	$0.515/GB	$2.58/GB	$3.61/GB	$12.88/GB
100 GB	$51	$258	$361	$1,288
512 GB	$263	$1,320	$1,848	$6,594
1,024 GB (1 TB)	$527	$2,641	$3,696	$13,189
5,120 GB (5 TB)	$2,636	$13,209	$18,483	$65,945
51,200 GB (50 TB)	$26,368	$132,096	$184,832	$659,456
102,400 GB (100 TB)	$52,736	$264,192	$369,664	$1,318,912

Excludes:
Curation
DMP Support
Tech Support

Challenges in Interpretation & Execution *(paraphrased)*

- “Charge secondary user only incremental cost”
- “Save reasonable data”
- “All… supporting digital data”
- “stored for long-term preservation and publicly accessible
- Breadth of NIH solution will differ by discipline: Clinical Trials – Genomic Data – Imaging – Basic Science
  - Some data need maximally for short period
  - Some longitudinal data gains value over time

Resources

- OSTP Public Access Policy Forum
  - https://www.whitehouse.gov/administration/eop/ostp/library/publicaccesspolicy
- NSF 16-009 - Public Access: Frequently Asked Questions
  - https://www.nsf.gov/pubs/2016/nsf16009/nsf16009.jsp
- The NIH Commons
  - https://datascience.nih.gov/commons
- National Institutes of Health Plan for Increasing Access to Scientific Publications and Digital Scientific Data from NIH Funded Scientific Research (February 2015)
  - http://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf
- ICPSR Data Management & Curation - Guidelines for OSTP Data Access Plan
  - https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/ostp.html#include
- Implementation of Public Access Programs in Federal Agencies
  - https://www.cendi.gov/projects/Public_Access_Plans_US_Fed_Agencies.html