Funding Agency Responses to Federal Requirements for Public Access to Research Results

by Wendy Kozlowski

Research Data Access & Preservation

EDITOR'S SUMMARY

A February 2013 government mandate directed federal departments and agencies to develop plans to promote access to the publications and data generated from federally funded scientific research. Panelists representing the National Oceanic and Atmospheric Administration (NOAA), the National Science Foundation's Directorate for Social, Behavioral and Economic Studies and the National Institutes of Health Office of Extramural Research shared updates on their agencies' progress. While each agency has more or less stringent policies in place for preserving data and making it accessible, the mandate expands their application without additional resources or guidance on writing or evaluating data management plans. The agencies represented will make efforts to minimize the burden on researchers, but audience members questioned who will handle the data management. Possible solutions include having librarians assigned to research projects or providing training to increase citation of datasets and link these outputs to grants.

KEYWORDS

government agencies

information policy

research data sets

data curation

access to resources

strategic planning

Wendy Kozlowski is science data and metadata librarian at the John M. Olin Library at Cornell University. She can be reached at wak57<at>cornell.edu.

On February 22, 2013, John P. Holdren, director of the federal Office of Science and Technology Policy (OSTP), released a memorandum for the heads of executive departments and agencies about increasing access to the results of federally funded scientific research. In the memo, the administration reiterated its commitment to ensuring, to the greatest extent possible, availability of federally funded scientific research results to the public, industry and the scientific community. The OSTP memo directs "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" [1]. It specifies a preference for agencies to work together to develop these plans and goes on to lay out objectives for access to both scientific publications and scientific data. Key drivers for this push toward increased access include maximization of "the impact and accountability of the Federal research investment" and enhancement of "innovation and competitiveness by maximizing the potential to create new business opportunities…" [1].

Affected agencies had until August 22, 2013 (six months) to submit a proposal. The OSTP, together with the Office of Management and Budget (OMB), reviewed and provided a first round of feedback on the drafts in February of 2014. Agencies had 90 days from that point to revise and resubmit final plans for review; it was in that window of time that this panel was held at the 2014 RDAP Summit.

The speakers on this panel shared updates on their agencies' efforts toward addressing the Public Access to Research Results (PARR) memo. The speakers were Jeff de La Beaujardière, data management architect at the National Oceanic and Atmospheric Administration (NOAA); Amy Friedlander,

staff associate in the office of the assistant director, National Science Foundation (NSF), directorate for social, behavioral and economic sciences; and Neil Thakur, special assistant to the National Institutes of Health (NIH) deputy director for extramural research. Neal Kaske, chief, central library and information services division of the NOAA central and regional libraries, joined the group for the question and answer portion of the session.

NOAA Plans for Improving Public Access to Science Research

De La Beaujardière began by describing the NOAA research domains as diverse and complex, with a wide variety of data types and formats; methods of data production, gathering and dissemination; and data collection purposes. Adding to the complexity, NOAA comprises five major units, each with established but sometimes incompatible and entrenched methods for data management (for example, the tornado warning system). Despite these challenges, interconnections and interdependencies among them exist, allowing NOAA to fulfill mission goals. Long before the OSTP memo was released, NOAA's vision was that all their environmental data should be discoverable, accessible and usable and preserved for all types of users across a variety of applications. The new federal mandate applies to NOAA-funded data and has provided guidance toward meeting these goals.

De La Beaujardière noted that there are two recent federal initiates involving open data: the 02-22-2013 OSTP PARR memo [1] and the 05-09-2013 OMB Open Data Policy memo [2]. Both apply to NOAA, and, while there is some overlap, the scope is different between the two, with the former affecting the grantees and the latter the NOAA employees and projects. Existing guidance for NOAA grantees requires data sharing in a timely fashion (no more than two years). In addition, NOAA projects must already plan how they will document, preserve and distribute their data, and NOAA requires ISO19115 metadata for discovery, use and understanding. New to NOAA will be sharing publications; they have a library, but NOAA repository efforts have focused on data archiving. NOAA's exact response to the PARR memo was still under review and could not be shared at the time of the panel, but de La Beaujardière did mention some potential impacts on their grantees, the largest of which will be in the area of data sharing.

Some of these possible outcomes include requiring data management plans (DMPs) from grant-issuing NOAA programs; requiring or encouraging use of the NOAA National Data Center for long-term preservation; more specific templates for grantee DMPs; inclusion of data sharing costs in proposal budgets (at the discretion of the grant-issuing agency); clarification or standardization of how funding sources are indicated in papers (for example, FundRef [3]); and adoption or establishment of methodology to ensure publications are broadly available after a suitable embargo period. Regardless of specific details, it was stressed that there would be an attempt to minimize additional effort on the part of the grantees, and that in general, feedback from OSTP/OMB has been positive for NOAA's proposed plan.

Finally, the NOAA presentation detailed recent data-related activities. De La Beaujardière talked first about the Dataset Identifier Project, which uses DataCite DOIs to link National Data Center data and metadata to publications (21 DOIs had been issued at the time of the RDAP conference, 66 at the time of preparation of this publication). Second, we learned about the NOAA Data Catalog beta-version website (http://data.noaa.gov), which is a CKAN-based catalog of datasets harvested from NOAA data centers (~48,000 datasets at the time of the RDAP conference, ~54,000 at the time of preparation of this publication). Finally, de La Beaujardière described NOAA's recent Big Data Partnership Request for Information, with the goal of increasing return on investment by using commercial enterprise to make NOAA's large body of datasets quickly available, at a large scale, to public users (see also [4]).

The NSF Public Access Initiative – Where Are We?

Amy Friedlander began by talking about not only the role of communication in maintaining transparency to taxpayers, but also its importance for progress in science and technology. Scholarly communication allows for vetting, validating and reproducing scientific findings, allows others to build on previous results and encourages innovation and the transition to products and services, a key goal of these recent data access initiatives. Friedlander went on to review the NSF's long-standing policy on data sharing, which includes the expectation that investigators will "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" [5, 6]. Building on this base, NSF began requiring data management plans (DMPs) as supplemental material for all proposals in 2011 [7], and those DMPs are evaluated as part of the merit review process. In 2013, the NSF began allowing datasets to be cited as relevant work products in biographical sketches [8] independent of the publication(s) based on them. In addition, the NSF allows request of funds to support publication and preparation of data for deposit as part of the budget proposal, and the http://nsf.gov and http://research.gov sites allow the public to search awards, research descriptions and publications (but not data).

Friedlander went on to describe the NSF's approach to public access as one that must consider the needs of researchers and investigators that study a broad diversity of science, come from a range of institutions, work with various publishers and often have the support of multiple funding streams. This diversity inherently impacts researcher behavior, and the NSF will deploy its plan in phases, allowing ongoing innovation as implementation occurs. As with NOAA, plan details could not yet be discussed, but the initial focus will be on sharing publications. NSF currently has no internal repository for content, nor have they shared plans to build one, so expectations are that content will go in discipline repositories. This practice will require interaction with other federal agencies and the research communities, as well as integration of internal systems within the enterprise architecture.

Looking forward, NSF expects to make no large changes to their existing requirements and practices because of the PARR memo. Once approved, the NSF will post its plan to the http://nsf.gov website, and any changes to current procedures will be announced with nine-month windows for notice and comment. More specific guidance may also be released at the program, division or directorate levels.

Public Access: NIH's Update

Keeping with the pattern of including a description of the agency's current open access efforts, Neil Thakur began his talk with a brief review of the NIH data sharing policy that has been in effect since 2003 [9]. These are among the key points: 1) data sharing plans are required for funding applications seeking $500,000 or more in direct costs in any year, or researchers must state why data sharing is not possible, 2) reviewers do not factor the proposed data-sharing plan into determination of scientific merit or priority scores and 3) specific program announcements may request data-sharing plans for proposals with less than $500,000 in direct costs.

NIH does not expect great changes to publications policy because of the PARR memo. Current requirements mandate that institutions and investigators are responsible for assuring copyright agreements are consistent with submission to PubMed Central (PMC). Upon acceptance for publication, authors are responsible for depositing the paper to PMC and then properly citing those articles with the PMC identification number in applications, proposals and reports as evidence of compliance. Thakur did touch on ways that institutions can help to ensure compliance with NIH policies, including training and author support for policy awareness, submission of manuscripts to PMC and proper preparation of citations. Also mentioned was offering support on understanding policies surrounding publishing agreements, communicating with publishers on behalf of researchers and the possibility that institutions might take a role in ensuring (or at least monitoring) institutional compliance.

Finally, Thakur stressed that preparation is key to avoiding funding delays. Authors need "plans that can withstand" forgetfulness and miscommunication among authors and between authors and publishers [10]. NIH encourages researchers to use My NCBI's My Bibliography service [11] to track their own compliance, associate papers with awards as soon as possible (don't wait until it's time to write a final report or apply for a renewal or new grant) and think about compliance plans as they write their papers, not at the last minute.

Commonalities, Differences, Impacts and Responses

Data management planning guidance is currently vague. Both de La Beaujardière and Friedlander acknowledged the lack of concrete guidance for writing and evaluating DMPs. NOAA discussed the possibility of

CONTENTS   TOP OF ARTICLE   < PREVIOUS PAGE   NEXT PAGE >   NEXT ARTICLE >

offering additional guidance as they move forward, but the NSF believes it is important for disciplines themselves to define what is important. Perhaps some combination of these two will shake out eventually, with communities helping to concretize best practices for planning. At least the initial stages of this process already seem to be happening within certain disciples, as seen in documentation prepared for the marine science community [12], the social science community [13], the ecology community [14] and others. Twitter discussion during the panel included whether or not this lack of specific guidance might actually lead to more anxiety on the part of grantees, as well as the reality that researchers do not necessarily know where to start designing DMPs [15].

There is a common desire to reduce impacts on researcher workload. Acknowledging the inevitability of some level of increased burden, all three agencies talked about the importance of making access to research results as least onerous as possible. Currently, NIH policy puts responsibility for publication upload to PMC in the hands of authors, but does offer several discipline-specific repositories for data. NSF discussed the possibility of using institutional and/or discipline repositories for research outputs, which, depending on implementation, may be a complicated and expensive alternative to (in the eyes of the researchers and institutions) agency-hosted repositories. NOAA has automated the move of project data from its repositories to its data catalog for increased discovery, but has not committed to that for grantee data, nor have they detailed how they will address sharing of publications. Given the top-down directive that plans be implemented using resources from within existing agency budgets [1], it will be interesting to see how well agencies are able to minimize administrative burden on researchers. Audience skepticism, based on ensuing twitter comments, seemed high, with specific concerns about library involvement in compliance monitoring and questioning where burden will actually fall ("if not the researchers, then who?") [15].

Libraries have a potential role in supporting researchers. Friedlander and Thakur mentioned several places that libraries might play a useful role in this migration to openly shared research results. Proper citation of publications, increased citation of datasets and linking of research outputs to grants were suggested as important places to focus. Everyone needs to consider who will actually do data management. One option is to increase embedded librarian efforts with researchers to provide or facilitate these activities. Other places that libraries are a likely fit are in training and outreach roles with the goal of increasing awareness and, when appropriate, researcher skills.

Despite the fact that panelists were not able to divulge details of their PARR memo plans, it was encouraging to hear that after over a year of waiting, there is indeed movement in this arena. Given the range of current positions and priorities of impacted agencies, it was also clear that there is no panacea for sharing research results. Nevertheless, it would appear that we can expect to see collaborative, iterative solutions that will evolve over time. As research and library professionals, one of our tasks will be to watch this space and communicate emerging requirements to the constituents we serve. ■

Resources Mentioned in the Article

[1] U.S. Office of Science and Technology Policy. (February 22, 2013). *Memorandum for the heads of executive departments and agencies. Subject: Increasing access to the results of federally funded scientific research.* Washington, DC: Executive Office of the President, Office of Science and Technology Policy. Retrieved from www.whitehouse.gov/sites/default/files/ microsites/ostp/ostp_public_access_memo_2013.pdf

[2] U.S. Office of Management and Budget. (May 9, 2013). *Memorandum for the heads of executive departments and agencies. Subject: Open data policy – Managing information as an asset.* Washington, DC: Executive Office of the President, Office of Management and Budget. Retrieved from www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf

[3] *FundRef*: www.crossref.org/fundref/

CONTENTS   TOP OF ARTICLE   < PREVIOUS PAGE   NEXT PAGE >   NEXT ARTICLE >

Resources Mentioned in the Article, cont.

[4] Hochmuth, C. (2014). NOAA embraces the buisness of big data. *FCW: The business of federal technology.* Retrieved from http://fcw.com/articles/2014/06/24/noaa-embraces-big-data.aspx

[5] National Science Foundation. (January 2010). *Award and Administration Guide.* Document number aag101, Section VI-D. Retrieved from
www.nsf.gov/pubs/policydocs/pappguide/nsf10_1/aag_index.jsp

[6] National Science Foundation. (February 2014). *Award and Administration Guide.* Document number aag14001, section VI-4. Retrieved from
www.nsf.gov/pubs/policydocs/pappguide/nsf14001/aag_index.jsp

[7] National Science Foundation. (January 2011). *Grant Proposal Guide: Significant Changes to the GPG.* Retrieved from
www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_sigchanges.jsp

[8] National Science Foundation. (January 2013). *Grant Proposal Guide.* Document number gpg13001. Retrieved from www.nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_index.jsp

[9] National Institutes of Health. (2003). *NIH Data Sharing Policy.* Retrieved from http://grants.nih.gov/grants/policy/data_sharing/

[10] Thakur, N. (2014). *Public Access: NIH's Update.* San Diego, CA: SlideShare. Retrieved from www.slideshare.net/asist_org/rdap-3-2714thakur?qid=f19bdaf1-6ae4-4ae5-a7f9-9d8f02f732a0&v=default&b=&from_search=10

[11] National Institutes of Health. (2014). *MyNCBI Sign In Page.* Bethesda, MD: National Center for Biotechnology Information. Retrieved from
www.ncbi.nlm.nih.gov/account/?back_url=http%3A%2F%2Fwww.ncbi.nlm.nih.gov%2Fmyncbi%2F

[12] Pollard, R.T., Moncoiffé, G., & O'Brien, T.D. (2011). *The IMBER data management cookbook - A project guide to good data practices.* IMBER Report No 3. Plouzané, France: IMBER IPO Secretariate, Institut Universitaire Européen de la Mer. Retrieved from www.imber.info/index.php/Science/Working-Groups/Data-Management/Cookbook

[13] Inter-university Consortium for Political and Social Research (ICPSR). (2012). *Guide to social science data preparation and archiving: Best practice throughout the data life cycle* (5th ed.). Ann Arbor, MI: Inter-university Consortium for Political and Social Research. Retrieved from www.icpsr.umich.edu/files/deposit/dataprep.pdf

[14] Strasser, C.A., Cook, R., Michener, W., & Budden, A. (2012). *DataONE primer on data management: What you always wanted to know but were afraid to ask.* Albuquerque, NM: DataONE. Retrieved from www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

[15] Briney, K. (2014). RDAP 2014 Panel – *"Funding agency responses to federal requirements for public access to research results."* San Francisco, CA: Storify.com. Retrieved from https://storify.com/KristinBriney/rdap-2014-funding-agency-panel