INTERAGENCY PUBLIC ACCESS COORDINATION

A REPORT TO CONGRESS ON THE COORDINATION OF POLICIES RELATED TO THE DISSEMINATION AND LONG-TERM STEWARDSHIP OF THE RESULTS OF FEDERALLY FUNDED SCIENTIFIC RESEARCH

Executive Office of the President
National Science and Technology Council

MARCH 2012

This report is submitted in fulfillment of Section 103 of the America COMPETES Reauthorization Act of 2010 (Public Law 111-358).

Introduction

The America COMPETES Reauthorization Act of 2010 (ACRA; Public Law 111-358), signed into law by President Obama in 2011, calls for the Director of the Office of Science and Technology Policy (OSTP) to establish a working group under the National Science and Technology Council (NSTC) with the responsibility to coordinate Federal science agency research and policies related to the dissemination and long-term stewardship of the results of unclassified research, including digital data and peer-reviewed scholarly publications, supported wholly or in part by funding from the Federal science agencies. This report to Congress—submitted in fulfillment of Section 103 of ACRA (Appendix VI of this document)—details progress toward the coordination of policies related to these goals.

The Administration has long recognized the importance of improving the management of and access to the results of federally funded scientific research including digital data and peer-reviewed publications. Since 2008, OSTP has been working to coordinate with agencies to develop policies that assure widespread public access to and long-term stewardship of the results of federally funded unclassified research. In 2009, the Interagency Working Group on Digital Data (IWGDD) under the NSTC Committee on Science (CoS) issued a report, entitled *Harnessing the Power of Digital Data for Science and Society*,[i] which called for improved management of digital data resulting from federally funded research to better leverage taxpayer investment in scientific research and development. That same year, OSTP issued an initial Request for Information (RFI) on access to scholarly publications resulting from research conducted with Federal funding to explore the need for and potential methods for increasing access to peer-reviewed scientific publications describing the results of federally funded research. That RFI showed broad support for increasing public access to scientific publications but disagreement on how best to achieve increased access.

Interagency Working Groups

The NSTC is a cabinet-level body that functions as the principal means within the executive branch to coordinate science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to establish clear national goals for Federal science and technology investments in a broad array of areas spanning virtually all the mission areas of the executive branch. Working groups of the NSTC are formed to expedite communication among the key agencies involved in funding and regulating federally funded scientific research. The work of the NSTC is organized under five primary committees: Environment; Natural Resources and Sustainability; Homeland and National Security; Science, Technology, Engineering, and Math (STEM) Education; Science; and Technology. Each of these committees oversees subcommittees and working groups focused on different aspects of science and technology and the necessary coordination across the Federal government.

ACRA requires that OSTP, through the NSTC, coordinate the development of Federal science agency policies related to the dissemination and long-term stewardship of the results of unclassified research, including digital data and peer-reviewed scholarly publications, supported wholly or in part by funding from the Federal science agencies.

These are necessarily deliberative processes among government officials. Sometimes these deliberations are informed through public comment such as the 3 RFIs issued on public access to peer reviewed scientific publications and the management of digital data resulting from federally funded scientific research. Using RFIs for public input into the NSTC deliberative processes ensured an equal playing field for all stakeholders.

Consistent with Section 103 (b)8 of ACRA, which requires a clear distinction between scholarly publications and digital data, two interagency groups were formed under the NSTC: the Task Force on Public Access to Scholarly Publications (PASP) and the Interagency Working Group on Digital Data (IWGDD). In turn, two subgroups are now functioning under the IWGDD, one focused on policies for preservation, discoverability, and access, and the other on standards for interoperability, re-use, and re-purposing. This was done because it became clear that any group working on interagency coordination of scientific standards would necessarily require participants with highly technical data management backgrounds, as well as experience in developing data policy requirements. Collectively, the deliberations of these groups will guide changes common to all science funding agencies on public access to peer reviewed scholarly publications and data management.

The interagency groups are tasked with identifying the specific objectives and public interests that need to be addressed by any policies in these two areas. They will take into account the varying missions, types of data, and dissemination models associated with the range of Federal science funded by different agencies and scientific disciplines. They will also help address other public access requirements of ACRA, keeping in mind the need to follow statutory requirements and best practices for protecting personal privacy, proprietary interests, intellectual property rights, and attribution. The objectives will be brought before the CoS, where agency leaders will consider next steps in the coordination of new interagency and agency-specific policies.

It should be noted that while ACRA defines the term "Federal science agency" as any Federal agency with an annual extramural research expenditure of over $100 million, we concluded that several science agencies that fall below that threshold nonetheless play a significant role in driving the U.S. scientific enterprise. Therefore several of these agencies have been participating in both formal and informal discussion with the NSTC groups. It should also be stressed that it is the intention of the Administration to continue a robust dialog with the private sector and the public to ensure that policies developed will benefit the public interest and to maintain a level playing field for all interested parties.

Interagency Working Group on Digital Data

The IWGDD existed previously, but was re-chartered under the CoS and explicitly tasked with developing options for implementing the digital data policy and standards requirements of Section 103 of ACRA and implementing the recommendations developed by the previous iteration of the IWGDD (Appendix V).

As noted, the new IWGDD consists of two groups, one focused on policies for preservation, discoverability, and access; and the other on standards for interoperability, re-use, and re-purposing. The 2009 IWGDD report, *Harnessing the Power of Digital Data for Science and Society*, provided a set of first principles that are guiding the vision, strategy, tactical goals, and implementation plans for the Federal government, acting as both leader and partner, to work with all sectors of our society to enable reliable and effective digital data preservation and access. The report specifically recommended that:

1. an NSTC Subcommittee for digital scientific data preservation, access, and interoperability be created;
2. appropriate departments and agencies lay the foundations for agency digital scientific data policy and make the policy publicly available; and
3. agencies promote a data management planning process for projects that generate preservation data.

Here, preservation data was defined as those data for which the benefits of preservation exceed the costs. This construct was designed to accommodate the varying standards within different scientific disciplines and communities of practice.

The current IWGDD includes representatives from the Department of Agriculture, Department of Commerce, Department of Defense, Department of Energy, Department of Health and Human Services, Department of Homeland Security, Department of State, Department of Transportation, Environmental Protection Agency, National Aeronautics and Space Administration, National Science Foundation, Department of Education, Department of the Interior, Department of Veterans Affairs, Institute of Museum and Library Services, Library of Congress, National Archives and Records Administration, the Smithsonian Institution, Office of Management and Budget, and OSTP.

The IWGDD is considering steps that can be taken by Federal agencies to encourage and coordinate the development of agency policies and standards to promote long-term preservation of and access to digital data resulting from federally funded research. This would include scientific data that are produced or collected by agency employees themselves as well as data that are produced by extramural researchers funded by Federal science agencies. We anticipate allowing considerable flexibility for agencies to

identify the research and data that fall under coordinated policies to allow for differing practices and standards in different scientific communities and to accommodate different agency missions.

It is expected that any agency policy for increasing accountability and access to digital data should necessarily follow best practices for protecting confidentiality, personal privacy, proprietary interests, legitimate intellectual property claims, and appropriate attribution, and for ensuring that homeland and national security are not compromised. Also, it is preferable that, when appropriate, scientifically valuable data resulting from federally funded research be deposited in publicly accessible databases if and when such repositories are available, and when the standard practices of a particular discipline make such deposition the norm. Such access to data will enhance discoveries and innovation, as well as public trust in science through verification and validation.

We anticipate the IWGDD will serve a critical role in coordinating the development of standards for scientific data to maximize interoperability across Federal agencies, across science and engineering disciplines, and between research data and scholarly publications. This will necessitate taking into account existing consensus standards, including international standards. This work will evolve over a period of years with different agencies taking the lead on encouraging the development of standards by joining forces with and helping to form communities of practice including private-sector and international partners.

Task Force on Public Access to Scholarly Publications

The PASP (Appendix IV) was established under the CoS. The PASP includes representatives from the Department of Agriculture, Department of Commerce, Department of Defense, Department of Energy, Department of Health and Human Services, Department of Veterans Affairs, Environmental Protection Agency, National Aeronautics and Space Administration, National Science Foundation, and the Executive Office of the President including the Office of Management and Budget and the OSTP.

The PASP is undertaking the scholarly publications portion of Section 103 of ACRA, with the responsibility to coordinate Federal policies related to the dissemination and long-term stewardship of peer-reviewed scholarly publications resulting from federally-supported unclassified research. The PASP continues to work on common objectives for the development of individual agency policies for ensuring public access to the results of federally funded research, including peer-reviewed scholarly journal articles and other peer-reviewed publications.

The PASP has gathered preliminary information, which includes inputs from a public consultation in 2009-2010 through an RFI; a report from the congressionally convened Scholarly Publishing Roundtable; and recent recommendations from associations, societies, companies, and other organizations through a second RFI issued in November 2011. These inputs have indicated a strong support for broad public access to scholarly publications resulting from federally-supported research. Similarly, agencies and public commenters are cognizant of the essential role that publishers and the peer review system play in advancing the scientific enterprise. The PASP therefore set out to explore what steps could be taken to expand public access while preserving the value that publishers provide to the scientific enterprise, creating new business opportunities, and maximizing the economic and societal benefits of the Federal investment in research and the resulting publications.

2009 Request for Information on Public Access Policies

In order to ensure that the voices of all stakeholders are heard equally, a series of RFIs was issued to collect ideas for how best to facilitate data sharing and public access to data and peer-reviewed scientific publications. In 2009, an initial public consultation about policy options for expanding public access to federally funded peer-reviewed scholarly articles was conducted through an RFI titled Public Access Policies for Science and Technology Funding Agencies Across the Federal Government (Appendix I).

We received more than 500 comments from a broad range of stakeholders including researchers, industry groups, publishers, and universities. Responses showed that the majority of stakeholders support increased public access to federally funded data, but differ on the form public access policy should take and how it should be implemented. Results of the 2009 RFI are available online.[ii]

2011 Requests for Information on Data Sharing and Public Access for Scholarly Publications

In 2011, two RFIs were released soliciting public input on long-term preservation of and public access to the results of federally funded research, including digital data (Appendix II) and peer-reviewed scholarly publications (Appendix III).

While the 2009 RFI sought public comment on policy options for expanding public access to federally-funded peer-review scholarly articles, these new RFIs took the process a step further, seeking guidance on access to scientific publications and initiating a parallel process relating to digital data, as called for in ACRA.

One hundred and eighteen comments on public access to digital data and 377 on public access to scholarly publications were received. These comments came from organizations and individuals representing a wide variety of fields and stakeholders including over 100 scientists, 30 publishers, 62 librarians, 55 scientific societies, and 17 voices from industry.

Results from the scholarly publications RFI showed that 89 percent of respondents supported agency action to allow public access to scholarly publications, and 68 percent of respondents who commented on an appropriate embargo period length (Appendix III, Question 8) were in favor of making publications freely available to the public within 12 months of the publication date.

Results from the digital data RFI showed broad support for both the principles of data sharing and public access and for the creation of requirements by Federal funding agencies that would put these principles into practice. There was a clear desire by many commenters to ensure flexibility and to balance the burden of preserving and sharing digital data against their inherent value to the scientific enterprise and overall public good. Approximately 85percent of respondents indicated support for increasing access to data produced in the course of federally funded scientific research, and more than 70percent supported a requirement for funding proposals to include a data management plan.

More information on the 2011 RFIs is available online.[iii]

Agency Progress

Federal agencies are heavily involved in the work on public access, both through the interagency working groups and through independent agency policy actions to promote data sharing and develop the standards and infrastructure necessary for effective data management, as well as to increase public access to reports of research.

The National Institutes of Health

Since 2003, the National Institutes of Health's Data Sharing Policy has expected all investigator-initiated applications with direct costs greater than $500,000 in any single year to address data sharing. In addition, NIH requires more specific data management and data sharing requirements for specific types of projects, such as genome-wide association studies.

In 2007, the NIH Public Access Requirement was signed into law. The Consolidated Appropriations Act, 2008[iv] states:

> *"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: provided, that the NIH shall implement the public access policy in a manner consistent with copyright law."*

The following year, President Obama signed the Omnibus Appropriations Act of 2009 into law, making this public access requirement permanent.[v]

PubMed Central (PMC), a database of full-text journal articles at the National Library of Medicine, was designed as a permanent digital archive; the ongoing cost of storing and retrieving additional articles is negligible once the article is converted to PMC's archival format. NIH assigns unique identifiers to each paper in the PMC database, and requires that these numbers are included in all citations. Violators of this policy may be subject to delayed or withdrawn funding.

The mechanics of this policy are straightforward:

1. NIH awards fund institutions to conduct research. Compliance with the Public Access Policy is a term and condition of award.
2. NIH awards are used to produce peer-reviewed papers. NIH awards fund salary support to write papers and publications costs, such as page charges and open access fees.
3. The author, as the creator of the work, holds the copyright in the original paper. The author gives NIH a non-exclusive right to copyright to the original paper in PMC and may transfer to the publisher the balance of his rights, including an exclusive copyright for the final published version of the paper.

4. Authors of papers using NIH funds may publish in any journal they choose, provided they reserve a portion of their copyright to ensure their final peer-reviewed author manuscript is posted to PMC. Alternatively, authors may make arrangements for the publisher to post the paper to PMC.
5. Publishers can choose to not review or publish papers under the provisions of the NIH Public Access Policy.
6. Once a paper has been accepted for publication, the author can submit his or her final peer reviewed manuscript to PMC (or the publisher can start the process), or the publisher can submit the final published article to PMC directly.

This policy, and its subsequent fine tuning, has led to a dramatic increase in the number of NIH papers posted to PMC. Since 2008, NIH has been able to collect over 260,000 papers under the Policy. Overall, the compliance rate stands at 75 percent and continues to edge upward. This success is due to the combined efforts of NIH, its investigators and the voluntary support of publishers. Thousands of journals voluntarily submit peer-reviewed author manuscripts to PMC to assist authors in complying with the Public Access process. Several hundred journal publishers voluntarily deposit final published versions of articles in PMC automatically on behalf of their authors. Publishers representing about 1000 journals voluntarily submit the full content of their journals to PMC, regardless of whether the issue contains an article subject to the NIH Public Access Policy.[vi,vii]

The NIH policy significantly expanded public access to the results of federally funded biomedical research and to date, there has been no demonstrable harm to the business of publishing biomedical research. This is particularly important given previous concerns by some in the publishing industry that the 12-month delay period would cause serious financial damage to publishers and scientific societies. While the U.S. economy has suffered a significant downturn in the past several years, the Science, Technical, and Medical (STM) publishing industry appears strong, with increases in both the number and price of STM journals. For example, the NIH public access policy requirement took effect in 2008. However, from 2007 to 2011, the number of biological sciences and agriculture journals and medicine and health journals grew by 15 percent and 19 percent, respectively.[viii]

In the same time span, the average price of biology journals and health sciences journals increased 26 percent and 23 percent, respectively[ix]. Further, the International Association of STM Publishers announced an independent forecast of increases in the growth rate of the medical journal market in the coming years, from 4.5 percent in 2011 to 6.3 percent in 2014.[x]

Papers collected under the NIH Public Access policy are made public on PMC, NIH's database of full-text biomedical research articles. Over 2.4 million articles are now in PMC, and about 10 percent of these have been collected under the NIH Public Access policy. The rest are supplied by other research funders and publishers. Every weekday, one half million users access the database, retrieving over 1 million articles. Based on internet addresses, an estimated 25 percent of users are from universities, 17 percent are from companies, and 40 percent from the general public.

The National Science Foundation

In February 2011, the National Science Board (NSB) Task Force on Data Policy assembled a Statement of Principles related to data sharing and management to guide its work.[xi] In a subsequent public meeting, the NSB Task Force heard presentations from invited speakers and sponsored panel discussions to provide perspectives on open access publishing from a broad range of experts and stakeholders. The NSB Task Force issued a report in December 2011 on how the National Science Foundation (NSF) can more effectively use digital research data to meet its mission. The report noted the complexity of data policy issues and highlighted the critical role that agencies and research communities will play in developing standards for data management. Further, the report called on researchers to take action toward sharing and preserving data within their own fields:

> *"To address the challenges associated with increasing scale, scope, and complexity of data, each science and engineering research community should take the responsibility for determining its own standards and conventions for data stewardship and for coordination across the research enterprise. Funding agencies and stakeholder communities must partner together during data policy development so that recommendations can be implemented by each science and engineering research community."*

The NSB Task Force report recommended that NSF require that the data, methods, and techniques published in research papers be made available for the purpose of building upon Federal investments and for verifying published conclusions. To ensure attribution, the NSB Task Force called for data to be shared and for the development of persistent electronic identifiers on all data and methods, which enable automatic attribution of authors and award funding. The full NSB report is available online.[xii]

In January of 2011, the NSF affirmed its data management policy requirement, by requiring that proposals include a data management plan that describes how funded researchers will conform to NSF policy on the dissemination and sharing of research results. The NSF policy is clear: "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." More information on this policy is found here: http://www.nsf.gov/bfa/dias/policy/dmp.jsp. Submissions that fail to include a data management plan shall be rejected. Data management plans shall be reviewed using NSF's merit review mechanism and as such shall constitute part of the proposal review process. As of January 31, 2012 over 50,000 proposals with data management plans have been submitted. NSF maintains links to additional data sharing requirements and guidance from Directorates, Offices and Programs.[xiii]

The NSF has identified access to digital products of NSF-funded research as one of its priority goals for FY2012-FY2013. This priority goal is intended to increase opportunities for research and education through public access to high-value digital data. Specifically the performance goal states, "By September 30, 2013, NSF will have established policies for public access to high-value data and software in at least two data-intensive scientific domains."

The America COMPETES Act (ACA) of 2007, Section 7010, requires that research outcomes and citations of published documents resulting from research funded, in whole or in part, by NSF be made available

to the public in a timely manner and electronic format. Research.gov's Research Spending & Results service provides transparent and open access to award information by offering a complete picture of the award, including: award details, award abstract at the time of award, and citations of publications resulting from the award.

In response to the ACA requirement, NSF implemented a Project Outcomes Report (POR) requirement for all new awards made or existing awards that receive incremental or supplemental funding on or after January 4, 2010. The POR for the general public is a new report, written by Principal Investigators (PIs) specifically for the public, to provide insight into the outcomes of NSF-funded research. These reports are posted on Research.gov for public viewing exactly as submitted by the PI or a co-PI. While not a substitute for peer-reviewed scientific publications in content or value, they are a source of information to members of the general public who are interested in learning how taxpayer dollars expand the Nation's scientific and engineering knowledge. Since the reporting requirement went into effect, about 3,500 PORs have been posted on Research.gov. In time, the value that the public finds in PORs will become clearer. Principles, objectives, policies and some pilots related to public access to peer reviewed publications, resulting from fully or partially funded research by NSF, are currently being discussed.

Department of Energy

The Department of Energy (DOE) provides electronic public access to full-text technical reports emanating from national laboratory and grantee R&D activities. Products such as the DOE Information Bridge offer searchable access to roughly 300,000 technical reports hosted at laboratories or at DOE's central repository.

In addition to technical reports, metadata for 10,000 scholarly articles, on average, are collected annually and made web-accessible through DOE's central repository. About 15 percent of these citations provide public access to authors' manuscripts, and hyperlinks to final journal articles are included in metadata records.

DOE has worked with the publisher-driven consortium, CrossRef, other Federal agencies, and individual publishers to identify funding agencies for each scholarly article. Specifically, CrossRef has added a new standard metadata element relating to funding agencies to its comprehensive database of scholarly articles. Another new metadata element will identify the contracts or grants that enable each article.

In addition to its agency-specific public access tools, DOE serves as the operating agent for the interagency federated search engine Science.gov. Established in 2002, Science.gov provides one-stop searching of 50 databases and 200 million pages of R&D information across 14 Federal agencies. Among the databases searched by Science.gov are several that deal with scholarly publications, including NIH's PubMed and PubMed Central, AGRICOLA (Department of Agriculture's catalog of citations to agricultural literature), DOE's Information Bridge, and other prominent sources at Defense, NASA, EPA, Interior, NSF, Transportation, Commerce, and other agencies. As Science.gov enables interoperable search across agencies' distributed databases, it is positioned to perform enhanced cross-agency searching of more scholarly publications as they become publicly accessible.

Summary and Future Steps

To summarize, the Administration been working on issues related to the management of and access to the results of federally funded scientific research. In accordance with ACRA, OSTP established the Task Force on Public Access to Scholarly Publications and re-chartered the Interagency Working Group on Digital Data under the NSTC CoS. Those groups are evaluating objectives for increasing access to and improving the management of the results of federally funded scientific research.

Three RFI's have been issued, two on public access to scholarly publications and one on the management of digital data. Responses to those RFIs are being analyzed now, but initial results show strong public support for increasing access to scholarly publications describing the results of federally funded research and for improving scientific data management and access. The NSTC groups are continuing to consider the public comments received from the RFIs and how they should be incorporated into the objectives required by ACRA. Once they have finalized their decisions, the objectives of all three groups will be combined and presented to the CoS. There, agency leadership will consider implementation options. In addition, the CoS will help prioritize the remaining responsibilities as described in ACRA Section 103 including further public consultation and international outreach necessary for developing agency-specific policies.

Appendix I: 2009 Request for Information: Public Access Policies for Science and Technology Funding Agencies Across the Federal Government

Background

On his first day in office, the President issued a Memorandum on Transparency and Open Government that called for an "unprecedented level of openness in government" and the rapid disclosure of one of our Nation's great assets—information. Moreover, the Administration is dedicated to maximizing the return on Federal investments made in R&D. Consistent with this policy, the Administration is exploring ways to leverage Federal investments to increase access to information that promises to stimulate scientific and technological innovation and competitiveness. The results of government-funded research can take many forms, including data sets, technical reports, and peer-reviewed scholarly publications, among others. This RFI focuses on approaches that would enhance the public's access to scholarly publications resulting from research conducted by employees of a Federal agency or from research funded by a Federal agency.

Increasing public access to scholarly publications resulting from federally funded research may enhance the return on federal investment in research in the following ways:

a. More timely, easier, and less costly access to scholarly publications resulting from federally funded research for commercial and noncommercial scientists has the potential to promote advances in science and technology, thereby enhancing the return on Federal investment in research;

b. Creating an easily searchable permanent electronic archive of scholarly publications resulting from federally funded research has the potential to allow cross-referencing, continuous long-term access, and retrieval of information whose initial value may only be theoretical, but may eventually have important applications;

c. Ensuring that the Federal agencies that support this research can access the published results has the potential to promote improved cross-government coordination of government funding, and thus improved management of the Federal research investments;

d. More timely, easier, and less costly access to scholarly publications resulting from federally funded research for educators and students, and "end users" of research, such as clinicians, patients, farmers, engineers, and practitioners in virtually all sectors of the economy, has the potential to promote the diffusion of knowledge.

The Executive Branch is considering ways to enhance public access to peer reviewed papers arising from all Federal science and technology agencies. One potential model, implemented by the National

Institutes of Health (NIH) pursuant to Division G, Title II, Section 218 of Public Law 110-161 requires that all investigators funded by the NIH submit an electronic version of their final, peer-reviewed manuscript upon acceptance for publication no later than 12 months after the official date of publication. Articles collected under the NIH Public Access Policy are archived in PubMed Central and linked to related scientific information contained in other NIH databases.

The NIH model has a variety of features that can be evaluated, and there are other ways to offer the public enhanced access to peer-reviewed scholarly publications. The best models may be influenced by agency mission, the culture and rate of scientific development of the discipline, funding to develop archival capabilities, and research funding mechanisms.

Invitation To Comment

Input is welcome on any aspect of expanding public access to peer reviewed publications arising from Federal research. Questions that individuals may wish to address include, but are not limited to, the following (please respond to questions individually):

1. How do authors, primary and secondary publishers, libraries, universities, and the Federal government contribute to the development and dissemination of peer reviewed papers arising from Federal funds now, and how might this change under a public access policy?
2. What characteristics of a public access policy would best accommodate the needs and interests of authors, primary and secondary publishers, libraries, universities, the Federal government, users of scientific literature, and the public?
3. Who are the users of peer-reviewed publications arising from Federal research? How do they access and use these papers now, and how might they if these papers were more accessible? Would others use these papers if they were more accessible, and for what purpose?
4. How best could Federal agencies enhance public access to the peer-reviewed papers that arise from their research funds? What measures could agencies use to gauge whether there is increased return on Federal investment gained by expanded access?
5. What features does a public access policy need to have to ensure compliance?
6. What version of the paper should be made public under a public access policy (e.g., the author's peer reviewed manuscript or the final published version)? What are the relative advantages and disadvantages to different versions of a scientific paper?
7. At what point in time should peer-reviewed papers be made public via a public access policy relative to the date a publisher releases the final version? Are there empirical data to support an optimal length of time? Should the delay period be the same or vary for levels of access (e.g., final peer reviewed manuscript or final published article, access under fair use versus alternative license), for Federal agencies and scientific disciplines?
8. How should peer-reviewed papers arising from Federal investment be made publicly available? In what format should the data be submitted in order to make it easy to search, find, and retrieve

and to make it easy for others to link to it? Are there existing digital standards for archiving and interoperability to maximize public benefit? How are these anticipated to change?

9. Access demands not only availability, but also meaningful usability. How can the Federal government make its collections of peer-reviewed papers more useful to the American public? By what metrics (e.g., number of articles or visitors) should the Federal government measure success of its public access collections? What are the best examples of usability in the private sector (both domestic and international)? And, what makes them exceptional? Should those who access papers be given the opportunity to comment or provide feedback?

Appendix II: 2011 Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research

In accordance with Section 103(b)(6) of the America COMPETES Reauthorization Act of 2010 (ACRA; Pub. L. 111-358), this Request for Information (RFI) offers the opportunity for interested individuals and organizations to provide recommendations on approaches for ensuring long-term stewardship and encouraging broad public access to unclassified digital data that result from federally funded scientific research. The public input provided through this Notice will inform deliberations of the National Science and Technology Council's Interagency Working Group on Digital Data.

Background

The multi-agency Interagency Working Group on Digital Data (Working Group), established under the National Science and Technology Council (NSTC) Committee on Science (CoS), has been tasked with developing options for implementing the digital data policy and standards requirements of Section 103 of ACRA. OSTP will issue a report to Congress, in accordance with Section 103(e) of ACRA, describing priorities for the development of agency policies for ensuring broad public access to the results of federally funded unclassified research, the status of agency policies for public access to digital data resulting from federally funded research, and a summary of public input collected from this RFI and other mechanisms.

The Working Group is considering steps that can be taken by Federal agencies to encourage and coordinate the development of agency policies and standards to promote long-term preservation of and access to digital data resulting from federally funded scientific research. Ideally, such policies would harmonize, to the extent practicable and feasible, data management plans for digital data that are collected or otherwise produced either by the agency itself or extramurally with Federal funds. The 2009 report of the Interagency Working Group on Digital Data of the National Science and Technology Council, "Harnessing the Power of Digital Data," recommended that agencies lay the foundations for digital scientific data policy and make their policies publicly available. It also recommended that agencies consider requiring data management plans for projects that will generate "preservation data"—those data for which the benefits of preservation exceed the costs. Federal science agencies already have some experience with policies to promote long-term preservation and access to scientific data.

Indeed current Federal policies promote and in many cases require Federal agencies to make the digital data generated by Federal agencies more publically accessible. However, such policies do not routinely cover data generated through Federal grants, cooperative agreements, and some other types of funding mechanism. Exceptions include, the National Institutes of Health's (NIH) Data Sharing Policy, which requires all investigator-initiated applications with direct costs greater than $500,000 in any single year provide a data management plan. In addition, NIH has more specific data management and data sharing requirements for specific types of projects, such as genome-wide association studies.

In January 2011, the National Science Foundation (NSF) reaffirmed its data management policy requirement, indicating that proposals must include a Data Management Plan that describes how funded researchers will conform to NSF policy on the dissemination and sharing of research results. The NSF policy is clear that "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." Such models may not necessarily be appropriate for all types of federally sponsored research.

As agencies consider how to further develop digital data policies, it is important to note that all policies for increasing accountability and access to digital data must follow statutory requirements and follow best practices for protecting confidentiality, personal privacy, proprietary interests, intellectual property rights, author attribution, and for ensuring that homeland and national security interests are not compromised.

The Working Group is now seeking additional insight from "non-Federal stakeholders, including the public, universities, nonprofit and for-profit publishers, libraries, federally funded and non-federally funded research scientists, and other organizations and institutions with an interest in long-term stewardship and improved public access to the results of federally funded research," as described in Section 103(b)(6) of ACRA. Specifically the Working Group seeks further public comment on the questions listed below:

Preservation, Discoverability, and Access

1. What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?
2. What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?
3. How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?
4. How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?
5. How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?
6. How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?
7. What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?

8. What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?

9. What mechanisms could be developed to assure that those who produced the data are given appropriate attribution and credit when secondary results are reported?

Standards for Interoperability, Re-Use and Re-Purposing

10. What digital data standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, Nature Genetics 29, 371) is an example of a community-driven data standards effort.

11. What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?

12. How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?

13. What policies, practices, and standards are needed to support linking between publications and associated data?

Response to this RFI is voluntary. Responders are free to address any or all the above items, as well as provide additional information that they think is relevant to developing policies consistent with increased preservation and dissemination of broadly useful digital data resulting from federally funded research. Please note that the Government will not pay for response preparation or for the use of any information contained in the response.

Appendix III: 2011 Request for Information: Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research

Purpose

In accordance with Section 103(b)(6) of the America COMPETES Reauthorization Act of 2010 (ACRA; Pub. L. 111-358), this Request for Information (RFI) offers the opportunity for interested individuals and organizations to provide recommendations on approaches for ensuring long-term stewardship and broad public access to the peer-reviewed scholarly publications that result from federally funded scientific research. The public input provided through this Notice will inform deliberations of the National Science and Technology Council's Task Force on Public Access to Scholarly Publications.

Background

The multi-agency Task Force on Public Access to Scholarly Publications (Task Force), established under the National Science and Technology Council (NSTC) Committee on Science (CoS), has been tasked with developing options for implementing the scholarly publications requirements of Section 103 of ACRA. OSTP will issue a report to Congress, in accordance with Section 103(e) of ACRA, describing priorities for the development of agency policies for ensuring broad public access to the results of federally funded unclassified research, the status of agency policies for public access to publications resulting from federally funded research, and a summary of public input collected from this RFI and other mechanisms.

In 2009 and 2010, OSTP conducted a public consultation about policy options for expanding public access to federally funded peer-reviewed scholarly articles. The Task Force has reviewed the information submitted through OSTP's public consultation (the full set of comments can be viewed on the OSTP website, experience with the various policies currently in use at a variety of Federal agencies, and a report from the congressionally convened Scholarly Publishing Roundtable.

The Task Force is now seeking additional insight from "non-Federal stakeholders, including the public, universities, nonprofit and for-profit publishers, libraries, federally funded and non-federally funded research scientists, and other organizations and institutions with a stake in long-term preservation and access to the results of federally funded research," as described in Section 103(b)(6) of the ACRA. Specifically, OSTP seeks further public comment on the questions listed below, on behalf of the Task Force:

1. Are there steps that agencies could take to grow existing and new markets related to the access and analysis of peer-reviewed publications that result from federally funded scientific research? How can policies for archiving publications and making them publically accessible be used to grow the economy and improve the productivity of the scientific enterprise? What

are the relative costs and benefits of such policies? What type of access to these publications is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise?

2. What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders involved with the publication and dissemination of peer-reviewed scholarly publications resulting from federally funded scientific research? Conversely, are there policies that should not be adopted with respect to public access to peer-reviewed scholarly publications so as not to undermine any intellectual property rights of publishers, scientists, Federal agencies, and other stakeholders?

3. What are the pros and cons of centralized and decentralized approaches to managing public access to peer-reviewed scholarly publications that result from federally funded research in terms of interoperability, search, development of analytic tools, and other scientific and commercial opportunities? Are there reasons why a Federal agency (or agencies) should maintain custody of all published content, and are there ways that the government can ensure long-term stewardship if content is distributed across multiple private sources?

4. Are there models or new ideas for public-private partnerships that take advantage of existing publisher archives and encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research?

5. What steps can be taken by Federal agencies, publishers, and/or scholarly and professional societies to encourage interoperable search, discovery, and analysis capacity across disciplines and archives? What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities? How should Federal agencies make certain that such minimum core metadata associated with peer-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to Federal science funding?

6. How can Federal agencies that fund science maximize the benefit of public access policies to U.S. taxpayers, and their investment in the peer-reviewed literature, while minimizing burden and costs for stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?

7. Besides scholarly journal articles, should other types of peer-reviewed publications resulting from federally funded research, such as book chapters and conference proceedings, be covered by these public access policies?

8. What is the appropriate embargo period after publication before the public is granted free access to the full content of peer-reviewed scholarly publications resulting from federally funded research? Please describe the empirical basis for the recommended embargo period. Analyses that weigh public and private benefits and account for external market factors, such as competition, price changes, library budgets, and other factors, will be particularly useful. Are there evidence-based arguments that can be made that the delay period should be different for specific disciplines or types of publications? Please identify any other items the Task Force might

consider for Federal policies related to public access to peer-reviewed scholarly publications resulting from federally supported research.

Response to this RFI is voluntary. Responders are free to address any or all the above items, as well as provide additional information that they think is relevant to developing policies consistent with increased public access to peer-reviewed scholarly publications resulting from federally funded research. Please note that the U.S. Government will not pay for response preparation or for the use of any information contained in the response.

Appendix IV: Charter of the Task Force on Public Access to Scholarly Publications

National Science and Technology Council

A. Official Designation

Pursuant to the requirements of Sec. 103 of the *America COMPETES Reauthorization Act of 2010* (COMPETES), the Task Force on Public Access to Scholarly Publications (PASP) is hereby established by action of the National Science and Technology Council (NSTC) Committee on Science (CoS).

B. Purpose and Scope

The purpose of the PASP is to develop recommendations for Federal policies related to the dissemination in peer-reviewed scholarly publications of the results of unclassified research supported wholly or in part by funding from the Federal science agencies, as defined In Section D below. To accomplish this, the PASP will:

1. Identify the specific objectives and public interests that need to be addressed;
2. Consider the existing diversity of research and dissemination models among Federal science agencies;
3. To the extent permitted by law, compile existing input and/or solicit new input from a variety of non-Federal stakeholders;
4. Consider the potential economic and other impacts to the science and engineering enterprise resulting from policies related to public access to peer-reviewed scholarly publications; and
5. Consider the role that scientific publishers play in the peer-review process in ensuring the integrity of the record of scientific research, including the investments and added value that publishers contribute.

C. Functions

The PASP will:

1. Develop and submit a draft set of recommendations for Federal policies on public access to the results of federally funded research in peer-reviewed scholarly publications;
2. Submit the recommendations to the CoS for consideration and, if deemed appropriate by the CoS, forward those recommendations to the Director of the Office of Science and Technology Policy (OSTP).

D. Membership

The COMPETES Act defines "Federal science agency" as "any Federal agency with an annual extramural research expenditure of over $100,000,000." The following eight departments and agencies meet this, criterion, as reported in the most recent National Science Foundation *"Survey of Federal Funds for Research and Development*. " Federal Funds for Research and Development: Fiscal Years 2007-09, NSF 10-302, May 2010.[xiv] These departments and agencies also predominantly support research that tends to result in unclassified peer-reviewed scholarly publications, and shall therefore be represented on the PASP:

Department of Agriculture;
Department of Commerce;
Department of Defense;
Department of Energy (Co-chair);
Department of Health and Human Services (Co-chair);
Environmental Protection Agency;
National Aeronautics and Space Administration; and
National Science Foundation (Co-chair).

The following organizations in the Executive Office of the President shall also be represented on the PASP:

Office of Management and Budget; and
Office of Science and Technology Policy.

Cooperating departments and agencies shall include such other Executive organizations, departments, and agencies as the Co-chairs of the PASP may designate.

E. Private-Sector Interface

The PASP may seek advice from members of the President's Council of Advisors on Science and Technology to secure appropriate private-sector advice, and will recommend to the CoS and/or the OSTP Director the nature of any additional private-sector advice needed to accomplish its mission. The PASP may also interact with and receive *ad hoc* advice from various private-sector groups consistent with the Federal Advisory Committee Act. The Federal Advisory Committee Act, 5 U.S.C. App., as *amended*, does not explicitly define "private sector," but the phrase is generally understood to include individuals or entities outside the Federal government such as, but not limited to, the following: non-Federal sources, academia, State, local or Tribal governments, individual citizens, the public, non-governmental organizations, industry associations, international bodies, etc. Additionally, the PASP may seek advice from the National Academies on the draft set of recommendations, pending availability of appropriated funds and consistent with law.

F. Termination Date

Unless extended by the Co-chairs of the CoS prior to its expiration, the PASP shall terminate no later than six months from the date of its first meeting or March 1, 2012, whichever occurs later.

G. Determination

I hereby determine that the establishment of the Public Access to Scholarly Publications Task Force is in the public interest in connection with the performance of duties imposed on the Executive Branch by law, and that such duties can best be performed through the advice and counsel of such a group.

Approved:

Francis Collins
Co-chair of the CoS, and
Director, National Institutes of Health

Subra Suresh
Co-chair of the CoS, and
Director, National Science Foundation

Carl Wieman
Co-chair of the CoS, and
Associate Director for Science
Office of Science and Technology Policy
Executive Office of the President

Appendix V: Charter of the Interagency Working Group on Digital Data

National Science and Technology Council

A. Official Designation

Pursuant to the requirements of Sec. 103 of the America Creating Opportunities to Meaningfully Promote Excellence in Technology, Education, and Science (COMPETES) Reauthorization Act of 2010 [P.L. 111-358], the Interagency Working Group on Digital Data (IWGDD) is hereby reestablished by action of the National Science and Technology Council (NSTC), Committee on Science (CoS).

B. Purpose and Scope

The purpose of the IWGDD is to develop recommendations for policies related to the preservation, access, and interoperability of scientific digital data produced through unclassified research supported wholly, or in part, by funding from the Federal science agencies. To accomplish this, the IWGDD will:

1. identify the specific objectives and public interests that need to be addressed;
2. take into account inherent variability among Federal science agencies, scientific disciplines, types of data, and preservation and access models;
3. coordinate the development or designation of standards for scientific data, the structure of full text and metadata, navigation tools, and other applications to maximize interoperability across Federal agencies, across science and engineering disciplines, and between research data and scholarly publications, taking into account existing consensus standards, including international standards;
4. in cooperation with the Networking and Information Technology Research and Development (NITRD) program, coordinate Federal science agency programs and activities that support research and education on tools and systems required to ensure preservation and stewardship of all forms of scientific digital data;
5. work with international science and technology counterparts to maximize interoperability between United States-based unclassified scientific data, tools, and repositories and their international counterparts;
6. in accordance with relevant law, solicit input and recommendations from, and collaborate with, non-Federal stakeholders, including the public, universities, libraries, federally funded and non-federally-funded research scientists, nonprofit and for-profit publishers, and other organizations and institutions with a stake in the long-term preservation of and access to the results of federally-funded research; and

7. coordinate Federal agency practices and procedures to promote preservation, discoverability, access, and interoperability to maximize opportunities for re-use and repurposing of scientific digital data.

C. Functions

The IWGDD will:

1. establish the following two task forces comprising member agency representatives with appropriate expertise:
   a. Task Force on Policies for Preservation, Discoverability, and Access; and
   b. Task Force on Standards for Interoperability, re-Use, and re-Purposing.
2. submit integrated recommendations from the task forces to the CoS for a review period not to exceed 30 days and incorporate the results of such review into the recommendations as appropriate; and
3. on or before October 1, 2011, submit the recommendations to the Director of the Office of Science and Technology Policy (OSTP) for consideration.

D. Membership

The COMPETES Act defines the term "Federal science agency" as any Federal agency with an annual extramural research expenditure of over $100,000,000. The following twelve NSTC departments and agencies meet this criterion, as reported in the most recent National Science Foundation "*Survey of Federal Funds for Research and Development*," and are therefore represented on the IWGDD:

Department of Agriculture;
Department of Commerce (Co-chair)
Department of Defense;
Department of Energy;
Department of Health and Human Services (Co-chair);
Department of Homeland Security;
Department of State;
Department of Transportation;
Environmental Protection Agency;
National Aeronautics and Space Administration; and
National Science Foundation (Co-chair).

The following departments and agencies have significant interests in digital scientific data preservation and access and are therefore also represented on the IWGDD:

Department of Education;
Department of the Interior;
Department of Veterans Affairs;
Institute of Museum and Library Services;

Library of Congress;
National Archives and Records Administration (Co-chair); and
The Smithsonian Institution.

The following organizations in the Executive Office of the President are also represented on the IWGDD:

Office of Management and Budget; and,
Office of Science and Technology Policy.

Cooperating departments and agencies shall include such other Executive organizations, departments, and agencies as the Co-chairs may designate.

E. Private-Sector Interface

The IWGDD may seek advice from members of the President's Council of Advisors on Science and Technology to secure appropriate private-sector advice, and will recommend to the CoS and/or the OSTP Director the nature of any additional private-sector advice needed to accomplish its mission. The IWGDD may also interact with and receive *ad hoc* advice from various private-sector groups as consistent with the Federal Advisory Committee Act.[xv]

F. Termination Date

Unless extended by the Co-chairs of the CoS prior to its expiration, the IWGDD shall terminate no later than April 1, 2012.

G. Determination

I hereby determine that the reestablishment of the Interagency Working Group on Digital Data is in the public interest in connection with the performance of duties imposed on the Executive Branch by law, and that such duties can best be performed through the advice and counsel of such a group.

Approved:

Francis Collins
Co-chair of the CoS, and
Director, National Institutes of Health

Subra Suresh
Co-chair of the CoS, and
Director, National Science Foundation

Carl Wieman
Co-chair of the CoS, and
Associate Director for Science
Office of Science and Technology Policy
Executive Office of the President

Appendix VI ACRA Sec 103

SEC. 103. INTERAGENCY PUBLIC ACCESS COMMITTEE.

(a) ESTABLISHMENT.—The Director shall establish a working group under the National Science and Technology Council with the responsibility to coordinate Federal science agency research and policies related to the dissemination and long-term stewardship of the results of unclassified research, including digital data and peer-reviewed scholarly publications, supported wholly, or in part, by funding from the Federal science agencies.

(b) RESPONSIBILITIES.—The working group shall—

(1) identify the specific objectives and public interests that need to be addressed by any policies coordinated under (a);

(2) take into account inherent variability among Federal science agencies and scientific disciplines in the nature of research, types of data, and dissemination models;

(3) coordinate the development or designation of standards for research data, the structure of full text and metadata, navigation tools, and other applications to maximize interoperability across Federal science agencies, across science and engineering disciplines, and between research data and scholarly publications, taking into account existing consensus standards, including international standards;

(4) coordinate Federal science agency programs and activities that support research and education on tools and systems required to ensure preservation and stewardship of all forms of digital research data, including scholarly publications;

(5) work with international science and technology counterparts to maximize interoperability between United States based unclassified research databases and international databases and repositories;

(6) solicit input and recommendations from, and collaborate with, non-Federal stakeholders, including the public, universities, nonprofit and for-profit publishers, libraries, federally funded and non federally funded research scientists, and other organizations and institutions with a stake in long term preservation and access to the results of federally funded research;

(7) establish priorities for coordinating the development of any Federal science agency policies related to public access to the results of federally funded research to maximize the benefits of such policies with respect to their potential economic or other impact on the science and engineering enterprise and the stakeholders thereof;

(8) take into consideration the distinction between scholarly publications and digital data;

(9) take into consideration the role that scientific publishers play in the peer review process in ensuring the integrity of the record of scientific research, including the investments and added value that they make; and

(10) examine Federal agency practices and procedures for providing research reports to the agencies charged with locating and preserving unclassified research.

(c) PATENT OR COPYRIGHT LAW.—Nothing in this section shall be construed to undermine any right under the provisions of title 17 or 35, United States Code.

(d) APPLICATION WITH EXISTING LAW.—Nothing defined in section (b) shall be construed to affect existing law with respect to Federal science agencies' policies related to public access.

(e) REPORT TO CONGRESS.—Not later than 1 year after the date of enactment of this Act, the Director shall transmit a report to Congress describing—

(1) the specific objectives and public interest identified under (b)(1);

(2) any priorities established under subsection (b)(7);

(3) the impact the policies described under (a) have had on the science and engineering enterprise and the stakeholders, including the financial impact on research budgets;

(4) the status of any Federal science agency policies related to public access to the results of federally funded research; and

(5) how any policies developed or being developed by Federal science agencies, as described in subsection (a), incorporate input from the non-Federal stakeholders described in subsection (b)(6).

(f) FEDERAL SCIENCE AGENCY DEFINED.—For the purposes of this section, the term "Federal science agency" means any Federal agency with an annual extramural research expenditure of over $100,000,000.

Endnotes

i. http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

ii. http://www.whitehouse.gov/administration/eop/ostp/library/publicaccesspolicy

iii. http://www.whitehouse.gov/blog/2012/01/30/your-comments-access-federally-funded-scientific-research-results

iv. http://www.gpo.gov/fdsys/pkg/BILLS-110hr2764enr/pdf/BILLS-110hr2764enr.pdf (PL 110-161, Division G, Title II, Section 218)

v. http://www.gpo.gov/fdsys/pkg/PLAW-111publ8/pdf/PLAW-111publ8.pdf (PL 111-8, Division F, Section 217)

vi. These publishers are listed under Method D at http://publicaccess.nih.gov/select_deposit_publishers.htm

vii. See http://publicaccess.nih.gov/submit_process_journals.htm for a full list

viii. http://ulrichsweb.serialssolutions.com/. Data accessed 11/18/11. Therefore, 2011 data may be undercounted.

ix. Library Journal Periodicals Price Surveys, 2009-2011. http://www.libraryjournal.com/article/CA6651248.html; http://www.libraryjournal.com/article/CA6725256.html

x. http://www.stm-assoc.org/industry-statistics/stm-subsegment-forecast-2011-2014/

xi. http://www.nsf.gov/nsb/committees/dp/principles.pdf

xii. http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf

xiii. http://www.nsf.gov/bfa/dias/policy/dmp.jsp

xiv. http://www.nsf.gov/statistics/nsf10305/pdf/nsf10305.pdf

xv. The Federal Advisory Committee Act, 5 U.S.C. App., *as amended,* does not explicitly define "private sector," but the phrase is generally understood to include individuals or entities outside the Federal government such as, but not limited to, the following: non-Federal sources, academia, State, local or Tribal governments, individual citizens, the public, non-governmental organizations, industry associations, international bodies, etc.