

Subject: Response to the OSTP RFI on Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research

Date: December 21, 2011 5:14:57 PM EST

December 21, 2011

White House Office of Science and Technology Policy
Request for Information on Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research.

Response from Arizona State University Libraries

(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 publicly 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?

For the purpose of this request for information, the primary step that agencies could take to grow existing and new markets related to access and analysis of peer-reviewed publications would be to develop policies that require free, unrestricted access to all articles resulting from federally funded research. Public access to this research will encourage and enhance scientific innovation and creativity by allowing researchers access to the most current research discoveries and enable them to leverage new information discovery technology and tools to make unforeseen connections, promote serendipitous discovery, and increase productivity. These policies can be used to grow the economy and improve the productivity of the scientific enterprise in a variety of ways. For example:

- Allowing broad and full reuse allows researchers to build on the results of others, rather than repeating research that has already been done. Researchers can also utilize new technological tools, such as applications that mine publication databases and produce meta-analyses of scientific results from multiple studies and datasets to make new connections and discoveries that would not be possible without computer assistance. Full open access to peer-reviewed publications encourages diversity, innovation, and efficiency, which improves the productivity of the scientific enterprise.
- Public access increases the pool of possible participants by taking advantage of unexpected readers who can innovate in ways not anticipated by the original researchers or their funders. These unanticipated users, who include commercial innovators, small companies, under-resourced researchers and potential collaborators, can hasten scientific discovery and innovation through engaging a larger pool of participants. For example, a 2006 study showed that sharing information about unsolved scientific problems to a large group of outside users was an effective and efficient means of finding solutions to problems that had not be solved internally (<http://www.hbs.edu/research/pdf/07-050.pdf>).
- Public access increases the global reach and impact of federally-funded research, thus enhancing the prestige of the U.S. scientific community.
- It can grow the economy through providing the opportunity for new information-based business models, as well as driving software development to facilitate new types of information discovery and tools for research.
- Faster commercialization spurs economic growth in broad sectors of the economy, such as

biotechnology, agriculture, and renewable energy, and encourages private investment in information technology to capitalize on government resources.

Some relative costs and benefits of such policies are as follows:

- Public access policies will increase the public's return on investment in research. It can be argued that the public currently pays multiple times during the research life cycle and may still not be able to access to the results. A public access policy could limit that to two (taxes that pay for the salaries and infrastructure for scholars and taxes that are used for large federal funding agencies), meaning that the public could immediately access the work once published without having to pay again. A relative cost will be reduced by a broader public access policy.
- Public access policies can reduce education costs by providing free access to authoritative scientific publications, while also engaging students and encouraging teachers to participate in the scientific community.
- Public access policies will increase government agency accountability and support informed, transparent, and evidence-based budget and policy decision-making by providing government agencies with an improved accounting of the outcomes of research they have sponsored, congressional budget drafters, appropriators, and authorizers with more reliable information to assess the value of existing expenditures and target funding priorities, and will promote more informed policy debate and enhanced access to the information underpinning policy decisions.
- An NIH-style public-access policy has proven to be cost effective: the NIH reports it costs \$3.5 - \$4.6 million annually (on a \$30 billion budget) to provide access to results of all their funded research. Therefore, an investment of about 1/100th of 1 percent of NIH's overall budget results in access to 2.2 million articles. As access and reuse is increased beyond the NIH model, that return on investment will rise even higher.

In conclusion, free, digital, online and immediate access is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise. Access to information without the ability to reuse or build upon it limits the value of that information.

(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?

The most important consideration regarding intellectual property protection lies with the researchers and authors, who are the initial rights holders and remain so until or unless they transfer their copyright to another, such as through publishing agreements. However, since most researchers are not paid for the production of scholarly publications, their primary concern with regards to copyright is its effect on their desire to have the greatest possible impact on their research field through attribution and citation; meaning they benefit most from the widest dissemination of their work. For this reason, public access does not threaten the incentives for the creation of scholarship at all - it enhances them because it supports faster research, more innovation and greater impact on the field.

Providing public access to federally-funded research requires no change in copyright law, but merely requires that the licenses that authors give to the agencies as a condition of their funding and the licenses that the agencies grant to the public facilitate both access and maximum reuse

of research information. A Creative Commons attribution license is an example of a license that would fulfill those purposes.

Agency policies should be mandatory for authors with a requirement to deposit their final (post-peer-review) author manuscripts, as this has proven to be successful for the NIH policy. Consequently, publisher agreements for federally-funded research articles should not be structured in ways that conflict with the licenses that researchers grant to the agencies. However, in order to ensure the continued value of the publisher's copy, we recommend that intellectual property policies require that the use of published scholarly works include a citation that refers to the publisher's version as the copy of record. This policy framework would balance the needs and interests of research authors, agencies, publishers, and the general public.

A copyright-related policy that should not be implemented is the current PubMed Central prohibition on systematic downloading, which needlessly restricts use, imposes unnecessary restrictions, and hinders scientific advancement.

(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?

A centralized approach to managing public access to federally funded research has many pros and very few cons. A centralized approach, managed by the federal government, would ensure that uniform standards for discovery, access, and persistence are maintained. Additionally, a centralized approach is more cost effective, both with regard to policy development and adherence, and in infrastructure development and maintenance. A standardized set of submission requirements facilitate policy compliance by making it simple for researchers to understand and follow a single set of workflows. As mentioned above, the NIH has proven that maintaining a central repository is very economical. This success could be expanded to include other federal agencies.

However, the most important aspect of public access to federally-funded research is the ability to search across a broad and interdisciplinary set of databases and have full and unrestricted access to the results. With this in mind, a decentralized approach involving multiple repositories might be a viable solution. In order to facilitate access and use of federally-funded research, federally-managed standards for access, interoperability, search functionality across repositories, adherence to standards, long-term archiving and preservation, openness and accountability, and the potential for creative reuse for research and commercial purposes would be required. It is important to consider that, even with carefully crafted regulatory requirements, it is more difficult to establish and maintain strict standards under a decentralized framework that includes partners outside the federal government, which is a considerable downside to a decentralized approach. Even if a decentralized approach proved feasible, it would be desirable for the federal government to maintain mirrored and accessible versions of any participating repositories in order to protect the public's investment and ensure accountability.

The federal government has a long-term interest in making such works permanently available, and is the only current entity with the interest in and capability for making federally funded works publicly available, in providing tools for use, and in guaranteeing that new services and

products can be built from publicly funded information. It is the only entity with the ability to establish and enforce standards of interoperability. The federal government is therefore the most appropriate body to maintain custody of publicly funded articles.

(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?

Academic and research institutions have already partnered with many research funders to provide permanent archives and have proven their expertise and focus on long-term preservation and access. The DRIVER project (<http://www.driver-repository.eu/>), funded by the European Commission, is an excellent example of a decentralized, federated repository structure involving cooperation from universities and research institutes in several European countries.

Publishers could be encouraged to participate in public-private partnerships by voluntarily providing the final published versions of articles. However, given their focus on immediate income, commercial publishing firms in particular should not be relied upon solely for digital archiving. To ensure long-term preservation, under no condition should a publisher site - or any other stakeholder site - be the single point of access for publicly funded articles.

As mentioned above, regardless of where partners are found, standards for access, preservation and interoperability must be maintained. Since a primary goal is to encourage innovation in accessibility and interoperability, it is important that creativity and experimentation be allowed in the tools that will be used in searching the publicly accessible articles. The ability to create new kinds of search tools, and to compute on the contents of these repositories (text-mining, etc.), must be afforded to users in order to exploit the full range of opportunities for innovative research and economic growth.

(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?

Federal agencies, publishers, and scholarly and professional societies can encourage interoperable search, discovery and analysis capacity across disciplines and archives by involving agencies that have experience in the development of metadata schema, such as NISO and the Library of Congress, in the creation of appropriate standards. Using previously developed models means that researchers and librarians, as well as automated systems, already have search techniques that work. At the minimum, core metadata standards for scholarly publications should be designed to facilitate the functions of use, reuse, and analysis described in previous responses.

Federal agencies can 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 by creating specifications that support multiple metadata standards. Some of the primary goals of the specifications (along with examples of related standards) would be to:

- Provide institutional information for published sources (grant IDs, funding organization, I2 – Institutional Identifier, etc.),
- Provide descriptive information for both the repository and published versions (Dublin Core, ORCID),
- Support searching through both key words and controlled vocabulary schema appropriate to disciplines,
- Incorporate abstracts,
- Facilitate full text searching and web crawling,
- Support metadata harvesting (OAI-PMH),
- Establish relationships through semantic web standards (RDF),
- Support usage tracking (COUNTER),
- Support description of related data (DataCite Metadata Schema),
- Support data exchange standards (JSON),
- Document IP rights.

(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?

Federal agencies that fund science can maximize the benefit of effective public-access policies to taxpayers by ensuring that complete results are widely available in a timely manner. To this end, any successful public-access policy must be consistent in its requirements and mandates. Researchers often hold grants from multiple agencies concurrently, so uniform requirements and procedures regarding deposit of peer-reviewed literature should be established across all funding agencies involved. In particular, a standardized set of criteria to determine what materials need to be deposited and how deposit should occur will reduce the complexity and cost and increase the rate of compliance. Such uniformity will also ease the burden on institutional and university repositories who often offer assistance to their researchers in meeting the deposit requirements for granting agencies such as NIH.

In order to minimize the burden and costs for stakeholders, federal policies should take advantage of existing protocols, such as SWORD, to facilitate automatic deposit of manuscripts to multiple repositories and should encourage development of additional tools and services. Additionally, licenses should be standardized to include the maximum possible reuse rights, such as those embodied in the Creative Commons CC-BY licenses, for the public.

Policies that integrate articles with grants management systems, both internal and external, will improve agency accountability, providing the public with more information about the results of their investment in scientific research. Policies can create opportunities to create enhanced productivity management tools for federal and internal reporting, such as enhanced bibliographies and primary investigator (PI) profiles. These tools would offer new opportunities for universities to better measure research output and promote branding of research.

It is important to note that these policy changes would address a burden that has been on the public for a long time - the lack of access to the research it funds. Without public access to federally-funded research, the public, small businesses, K-12 and higher education has borne extra cost burdens without the benefit of access to the products it funds.

(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?

The majority of scientific knowledge resulting from federal funding appears in form of peer-reviewed journal articles, the primary mechanism for scientific communication. As noted above, dissemination of the results of federally-funded research is hampered by limitations on access to journal literature – to the detriment of science, economic growth, and the general public interest. For those reasons, agency policies should concentrate their public access policies on journal articles as well as related supporting materials that document the research process, such as data, protocols, survey instruments, etc., and facilitate the replication of results.

Educational materials, such as book chapters, texts, and conference proceedings, resulting from publicly funded research should also be made accessible to the public. However, different conditions may apply to different types of materials. For example, authors are generally not paid for journals articles, but may be paid for textbook chapters. Any policies should reflect these differences.

In any case, the immediate effort should stay focused on the primary mechanisms that scientists use to communicate their results, namely peer-reviewed scholarly journal articles, and the current limitations on access to such publications. The complexity of providing access to other materials should not be allowed to delay the implementation of a mandate for public access to funded articles across federal agencies.

(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?

As mentioned in comments above, immediate access to the full content of peer-reviewed scholarly publications resulting from federally funded research is strongly recommended. Any embargo period diminishes the possibilities for innovation and growth that a mandate is trying to realize. In economic terms, each day of an embargo period represents lost opportunity costs that decrease the return on investment for the policy. For this reason, an embargo period is a compromise and should be as short as possible.

Empirical evidence supporting the value of public access is plentiful and growing all the time. Therefore, the burden of proof for the need for embargoes should be provided by those who desire them. For example, there is no evidence that libraries have cancelled journal subscriptions because of publicly accessible research articles. Before this and similar assertions are allowed to impose limitations on public access, evidence should be required to demonstrate actual harm. Where no such evidence can be provided, immediate public access should be the granted as the best way to foster innovation, competition, economic growth and scientific progress.

If it can be demonstrated that an embargo is necessary, a uniform embargo period of six months or less should apply across all funding agencies.

These comments are submitted on behalf of Arizona State University Libraries by:

Anali Maughan Perry
Collections and Scholarly Communications Librarian

Sherrie Schmidt
University Librarian