

Response to  
**Request for Information: Public Access to Peer-Reviewed Scholarly Publications  
Resulting From Federally Funded Research.**

Submitted by  
James L. Mullins, PhD  
Dean of Libraries and Professor  
Purdue University  
December 2011

## Background

Access to scholarly literature and other outputs of research is of great concern to Purdue University and the larger academic community. At Purdue, we strongly support White House action in support of long-term stewardship and broad public access to the scholarship resulting from federally funded research.

Purdue University is a coeducational, state-assisted system in Indiana. Founded in 1869 and named after benefactor John Purdue, we are one of the nation's leading research institutions with a reputation for excellent and affordable education. Our West Lafayette campus offers more than 200 majors for undergraduates, over 70 master's and doctoral programs, and professional degrees in pharmacy and veterinary medicine. The University expended \$472.7 million in support of research system-wide in 2006–07, using funds received from state and federal governments, industry, foundations, and individual donors. The West Lafayette campus boasts more than 400 research laboratories and 116 University-approved research centers and institutes. Purdue's Office of Engagement matches Purdue expertise and resources with the expressed needs of business, governmental agencies, communities, schools, and individuals around the world, with a specific focus on improving the economic prosperity and quality of life for Indiana's residents. (<http://www.purdue.edu>)

At Purdue, we believe that the commercial academic journal pricing situation makes the current system of scholarly communication unsustainable. Scholarly journals from commercial publishers are too numerous and increase in cost at a rate that outpaces inflation and the consumer price index. For several years academic libraries have been working together and with university presses to explore alternative means of publishing scholarly content, with fundamental goals to reduce costs and increase access. In 2006 we developed Purdue e-Pubs, Purdue's digital repository for Purdue-based scholarship. e-Pubs, like other open access repositories, into which journal articles and other scholarly content may be archived, comprise an important part of the Open Access equation, particularly for libraries and institutions looking to preserve and provide access to scholarly outputs originating from the home institution. Open Access repositories are enabled in large part by the willingness of authors to negotiate terms of publication in ways that permit deposit of publication into a digital archive, often after an appropriate "embargo" period during which publishers can recoup their investments.

With respect to the specific questions asked in the Request for Information, we recommend the following:

**(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?**

Federal agencies such as the Institute for Museum and Library Services have supported a number of important initiatives to grow existing and new markets, and to increase the productivity of the scientific enterprise. Continuing support for IMLS is essential. Two recent examples of relevant projects in which Purdue University has been involved demonstrate the positive effect that IMLS funding is having in increasing efficiency and improving productivity.

“Library Publishing Services: Strategies for Success” (<http://wp.sparc.arl.org/lps/>) was a research project conducted in 2010 and 2011 that brought national leaders involved in developing library-based publishing programs together for the first time. A survey conducted as part of the project revealed that almost 80% of Association of Research Libraries member institutions are either developing or implementing publishing services, and that libraries are increasingly taking leadership roles in publishing scientific research, particularly in professionalizing the dissemination of “gray literature” such as technical reports and conference proceedings that have previously been difficult to access.

Databib (<http://databib.lib.purdue.edu/about.html>) is at an earlier stage in development but will offer an online, community-driven, annotated bibliography of research data repositories which will, among other benefits, increase efficient use of existing repository infrastructure for preserving and disseminating scientific research and better leverage past federal and private funding.

**(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?**

Intellectual property interests include copyright, trademarks, and patents. For the purpose of this RFI we will address the copyright issue with the understanding that trademarks and patents involve different ownership issues that are specific to institutions.

This public access policy should not be restricted solely to scientific research, but should extend to include all federally funded research. Scholarly publications can range from journal articles to books to book chapters to conference proceedings, etc. For the purpose of this response our comments relate specifically to scholarly journal articles.

A variety of approaches are possible, from the most radical -- of all publicly funded research resulting in a scholarly journal article being considered a government work and immediately accessible, with no copyright attached to it, i.e.: public domain, to the more moderate approach where all funded research would be deposited into a freely accessible repository, with a reasonable embargo period such as six months to one year but no more than that, whether at a university, government agency, etc. Would it be reasonable that Library of Congress as our national library be one of the repositories for publicly funded research?

Fundamentally, a policy not recommended in any form would be to allow a copyright holder, whether publisher or author, to restrict access to the results of federally funded research for the duration of copyright.

**(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?**

For long term stewardship, the Library of Congress could play a role, in that the Library of Congress's mission is "to support the Congress in fulfilling its constitutional duties and to further the progress of knowledge and creativity for the benefit of the American people." The mission of LC's Library Services unit "is to develop qualitatively the Library's universal collections, which document the history and further the creativity of the American people and which record and contribute to the advancement of civilization and knowledge throughout the world, and to acquire, organize, provide access to, maintain, secure, and preserve these collections."  
(<http://www.loc.gov/about/>)

Pros and con exist for both centralized and decentralized approaches. With a decentralized approach there may be more quality control of content and organization or arrangement of content. With a centralized approach, the one-stop-shopping model to which most Internet searchers are accustomed to when searching the Internet may be more achievable.

**(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?**

HathiTrust, <http://www.hathitrust.org/>, a partnership of major research institutions and libraries could be considered one model of this type. HathiTrust states that it works to "ensure that the cultural record is preserved and accessible long into the future." Purdue is a member of HathiTrust and has made accessible through HathiTrust many Purdue-related publications.

The exponential increase in the amount of data generated by research in the digital age poses challenges to both publishers and libraries, and there are opportunities for joint initiatives that harness the expertise libraries have built in the organization and preservation of data to publishers' ability to effectively drive usage of it. Purdue University Libraries is one of the founding partners of the DataCite initiative, <http://www.datacite.org> that allocates stable digital identifiers to data so that it can be cited. The University has also been a leader in establishing a sustainable digital data repository, the Purdue University Research Repository (<http://research.hub.purdue.edu/>), the infrastructure behind which could be leveraged to create subject-specific collections of data in collaboration with specialist, especially learned society and university press, publishers.

There are as yet relatively few examples of this kind of library-publisher partnership to support a

disciplinary community, perhaps because society publishers in particular usually draw their members from multiple institutions and find it difficult to partner with one of them. However the work of the Digital Research and Curation Center (DRCC) at the Johns Hopkins University Sheridan Libraries in partnership with the American Astronomical Society (AAS) and its publishing partner, the Institute of Physics (IoP), has shown how digital archives, electronic publishing systems, and research communities can work together in a unified and efficient system.

**(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?**

Steps that can be taken to encourage search, discovery, and analysis include standardization of metadata for discovery; open sharing of core metadata; federal program or incentives to openly share citation data; standards development for the sharing of embedded and supplementary materials to enhance capacity for analysis. This would include usage of standard identifiers (e.g., DOIs) for such materials, but also the development of methods for embedding structured data formats from disciplines into electronic publications; and mandates for sharing of data (defined broadly) with publishers/societies more proactive about mandating data citation.

Minimum core metadata include metadata elements necessary for citation (see DataCite Metadata Kernel for an example) plus abstracts. Formal subject analysis would be nice, but is difficult and expensive to do in a cross-disciplinary manner. [The core metadata should include authors, title, date, source publication citation info \(ideally with persistent URL to access\) and keywords from controlled vocabularies \(not just discipline-specific, but also broader subject categories to be more widely shared across disciplines\) as well as source of funding \(agency\), this would allow linking to federal science funding.](#) While open metadata is important for search and discovery, open availability of full-text is necessary to allow disciplines to more fully engage the literature in their research (enables not only reading, but text mining apps).

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 standards development and core metadata, as described above; mandates that publishers share core metadata (including abstracts) openly (perhaps exposing them using OAI-PMH, or a similar protocol); and stronger mandates for sharing of full-text and data. [A wide variety of metadata standards are emerging, with Dublin Core, at least at this time, offering the most flexibility in terms of use by a wide variety of systems.](#)

**(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?**

While there are pros and cons to both centralized and decentralized approaches, the distributed and collaborative nature of 21st century scientific research and the networked character of digital information distribution suggest that no one organization or agency can bear the burden of preserving and delivering the results of taxpayer-funded research.

In this environment, initiatives that encourage the development of a range of repositories organized at either institutional or disciplinary levels and the support of a diverse publishing ecology, not dominated by a few large players, will maximize the benefit of public access policies. Existing disciplinary repositories such as arXiv (<http://arxiv.org/>) suggest how large “big science” communities can be served, although they struggle to find sustainability models. There is a need also to support “small science” in a way that is sensitive to the needs of a particular discipline, and some innovative partnerships between libraries and information technology services on campuses are starting to emerge to do this. Purdue University now supports over thirty “hubs” hosted on the HUBzero “platform for scientific collaboration” which was developed with substantial support from the National Science Foundation (<http://hubzero.org>).

In practical terms, federal agencies that fund science need to ensure that sufficient funds are included in grants to sustain the network of disciplinarily-sensitive repository services, need to continue to support the creation of standards for repositories and new forms of digital publication, and need to encourage the development of directories and “refer-itory” tools that help authors and users locate the partners best suited to preserve and disseminate their research findings.

**(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?**

We believe that all publications resulting from federally funded research should be made available freely to the public.

At Purdue, e-Pubs contains many publications that are not articles in scholarly journals, but provide much beneficial research to the public. These including technical reports, white papers, and conference proceedings. Use of these non-traditional research publications in Purdue’s e-Pubs is extremely high, as illustrated in a recent paper by Purdue University scholars accepted by the Transportation Research Board of the National Academies for its 2012 meeting, [http://docs.lib.purdue.edu/lib\\_research/146/](http://docs.lib.purdue.edu/lib_research/146/). The paper also suggests that investment in updated systems and infrastructure more well-g geared to digital distribution of so-called gray literature produced with State as well as Federal funding is essential and offers a model for innovative partnerships between public agencies and university libraries.

As the division between book and journal publications reduces, the exemption of book chapters from federal funding requirements for Open Access becomes less meaningful. It is true that book publication operates according to different norms and has traditionally involved more investment by the publisher, but most of the added value lies in the apparatus surrounding the articles (such as the index, introduction, part openers, consolidated bibliography, and organization) rather than the chapters themselves which are analogous to journal articles.

In general, the digital environment is removing traditional distinctions between types of publications that were developed in the print age. Any policy that discriminates between

publication formats based on their physical characteristics will soon become outmoded in the virtual environment and may be subject to abuse if less scrupulous commercial entities manipulate the formats in which research is clothed to impede access to publicly-funded scholarship.

**(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?**

The appropriate embargo period after publication is dependent on disciplinary norms but we believe that this need never be more than 12 months. A system that relies on calculating the “half life” normal to each discipline and sub-discipline introduces unnecessary complexity and would not be scalable. The empirical evidence is divided and in any case is based on snapshots of research practice that are unlikely to carry forward in a digital environment where the normal practices of research, and understandings of what is “current” and what is “archival,” are changing at an astonishing speed.

Other types of publications resulting from federally funded research, beyond traditional journal articles, should also be mandated for open access for the public because if publishers believe they are losing revenue they may make business decisions to do away with traditional scholarly journals and instead publish under the guise of another format (i.e. book) to avoid the mandate of deposit.