



Office of Science and Technology Policy (OSTP)  
[publicaccess@ostp.gov](mailto:publicaccess@ostp.gov)

January 11, 2012

On behalf of the University of North Carolina at Chapel Hill Libraries, I submit the following response to the request for information regarding "Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research."

We believe that free, open access to published research results is of overwhelming benefit to society. It should be the right of taxpayers to freely access findings from federally funded research paid for by tax dollars. Specific recommendations as requested in the ROI follow, with numbering referring to that document.

**#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?**

There is ample evidence that free public access to research results removes barriers and spurs innovation, to the benefit of society and the economy. The editor-in-chief of PLoS Computational Biology told recently of a ground-breaking manuscript he had received. The author was a 15-year-old high school student who had used a "test" login obtained from a vendor to access resources at a nearby university library. While not all stories may be so dramatic, the story illustrates the potential for students, faculty, and others to make transformational discoveries when the cost barrier of access to taxpayer-funded research is removed.

Open access has already led to new products and approaches within the publishing and information industries. To support these trends, federal agencies need to follow the lead of NIH and ensure that all federally sponsored research is publicly accessible free of charge. The best way to accomplish this is via agency-supported repositories like PubMedCentral. It is alarming to note that the recently introduced Research Works Act (HR 3699) would undermine this successful effort, and thereby substantially weaken our nation's capacity to assert leadership in science and innovation worldwide. We would be better served by policies that encourage both commercial and not-for-profit to adopt new business models that provide revenue without stifling overall scientific communication.

There is already evidence that some publishers are doing exactly that. The number of "gold" open access journals has grown rapidly, numbering 7,311 at the end of November, 2011 ([http://oad.simmons.edu/oadwiki/OA\\_by\\_the\\_numbers](http://oad.simmons.edu/oadwiki/OA_by_the_numbers)). This is a clear market indicator. Examples of new products and services that have built upon publicly accessible scholarship include GoPubMed (<http://www.gopubmed.org/>) and Google Scholar (<http://scholar.google.com/>). Such tools and products add value by abstracting, mining, synthesizing, and otherwise using freely available content.

Publishers who embrace the open access model, such as BioMed Central and Public Library of Science, have developed viable business models in order to make content freely available to all. Traditional publishers have adapted their business models too, using hybrid and author-pay options that protect their revenue base while achieving widespread public access. However, these latter approaches simply shift the cost from the publisher to the author, and in many cases are unaffordable by authors. Federal agencies should support these approaches by encouraging grantees to include open access publishing costs in their grant budgets. Such a policy identifies and supports publication costs as essential to the completion of the research process.

The economic benefit to society from such efforts is hard to calculate but could be enormous. There is ample evidence that freely accessible articles have considerably more use than fee-based articles. Reputable studies detail widespread economic and social benefits to society of free and open access to the world's peer-reviewed research and counter critiques from publishers (see, e.g., John Houghton and Charles Oppenheim, "The economic implications of alternative publishing models", *Prometheus*, March 2010, <http://dx.doi.org/10.1080/08109021003676359>).

One thing that federal agencies must do is expand on the current requirement to deposit articles in a publicly accessible repository like PubMedCentral. While institutions and publishers may also wish to serve as repositories, it is in the public's best interest for the federal government to guarantee permanent access to research articles either via its own repositories or by requiring others to maintain policies and standards for interoperability, accessibility, and permanent free access.

Scientific research is increasingly interdisciplinary. Productivity and benefit to society is enhanced by collaborations among organizations, public and private, that are engaged in translational research and development. (One example from North Carolina's Research Triangle Park is the Hamner Institute, of which a major unit is the Institute for Drug Safety Science.) This is a national priority as reflected in the Clinical and Translational Science Awards Program of NIH. These collaborations are not well served by the current cost model from most STM publishers, who charge high license fees typically based on institutional affiliation. This means that collaborators working side by side but with different affiliations, cannot access the same content. Our library has seen ample evidence of this via requests for access to our collections that we cannot honor. Public access for free, regardless of affiliation, helps promote such collaborations that can lead to new discoveries that improve the public's health. Free public archives of scholarship is equally important. The "translation" process often involves developing new commercial products via collaborative efforts and agencies such as the Hamner Institute. It also involves workforce training that leads to more jobs for skilled workers.

**#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?**

As librarians we recognize the need to balance the right of content creators with the rights of the general public that can benefit from the results of their research, as supported by the Copyright Act. However, the balance shifts when the research is taxpayer supported. We support the recent position at Harvard calling for all research funders to require the use of open licenses ([http://cyber.law.harvard.edu/files/OCL\\_for\\_Foundations\\_REPORT.pdf](http://cyber.law.harvard.edu/files/OCL_for_Foundations_REPORT.pdf)). One open license, the Creative Commons Attribution license (CC-BY) is a good model that maintains this balance by allowing

works to be available and usable with proper attribution to the author. It fully protects the author's intellectual property rights and fair use, which we believe is essential.

The current NIH policy also serves as a good model when considering this question. Authors now sign publication agreements with publishers retaining the author rights that allow them to deposit their articles in the NIH repository, PubMedCentral, while assigning other rights to the publishers.

**#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?**

PubMedCentral functions as a successful example of a centralized federal public access repository. It accomplishes the important functions of free, permanent access, long term preservation, and interoperability. Central repositories like PubMedCentral have these advantages:

- Systematic and comprehensive collection, less risk of inadvertent omissions;
- Consistency in collection policy and consistency in how the policy is applied to the work;
- A single source for researchers to find works rather than needing federated search tools to link multiple sources;
- Interdisciplinary searching for materials may be easier;
- Administration costs are likely to be lower. (NLM reports that PubMedCentral costs less than 1/100th of one percent of NIH's operating budget to run).

Central federal repositories signal a role for the federal government that is increasingly important now that most research publications are digital, not print. The role of research libraries to provide long term access to print content is not easily transferred to the digital world in which those libraries rarely actually own the digital content, but merely license it. However, it still should be a goal to achieve redundancy in digital repositories, whether they are maintained by libraries, publishers, scholarly societies, consortia, or others.

There are advantages to decentralized repositories, too. Redundancy is needed to guarantee permanent access to needed content. Separate repositories allow for focusing on different disciplines and circles of interest. Such repositories can be designed to meet the needs of the specific types of materials likely to be collected for a specific discipline. In the long run, society will be best served by the federal government taking a leadership role and working collaboratively with local partners.

**#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?**

Public/private partnerships that take advantage of existing publisher archives should be encouraged as long as they are "trusted" repositories that meet all the conditions for public accessibility, use rights, interoperability, and the long-term preservation of publicly funded articles. Emphasis should be given to the fact that no publisher or any other single stakeholder be the single point of preservation and access for these articles.

Research universities and their associated libraries have extensive experience in preservation, archive infrastructure, and access. In fact, the best public-private partnership model (specifically Federal-university) is the successful ArXiv.org that provides open access to 727,246 e-prints in the fields of physics, mathematics, computer science, quantitative biology, quantitative finance, and statistics. Another emerging model is the HathiTrust, “a partnership of major research institutions and libraries working to ensure that the cultural record is preserved and accessible long into the future.” The Triangle Research Libraries Network (UNC-Chapel Hill, Duke, NC State, and NCCU) are community partners in this effort.

**#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?**

Open metadata is key to facilitating the wide dissemination and use of scholarly works, including those funded by federal research dollars. While services that promote discovery and re-use of scholarly work increasingly use the full text of these works in their operation, metadata is still an essential part of the resource sharing ecosystem. Structured metadata provides necessary context that supplements the information in textual works, and is critical in discovery and re-use of datasets, images, video, and other types of scholarly work that are not textual in nature.

There is significant evidence that open sharing of metadata facilitates use of research and digital content. The Europeana project recently issued a white paper that lays out a business case for its metadata to be issued under an open license, which can be found at <http://www.scribd.com/doc/73652620/Europeana-White-Paper-2>. The Library of Congress’ posting of a sampling of their photographic holdings, with metadata, on Flickr (in addition to launching the Flickr Commons) significantly increased user interaction with this material. In the first day, Library of Congress images on the Flickr Commons received 1.1 million page views, and in the first week they received 3.6 million page views (<http://blogs.loc.gov/loc/2008/01/flickr-followup/>, [http://www.loc.gov/rr/print/flickr\\_report\\_final.pdf](http://www.loc.gov/rr/print/flickr_report_final.pdf)). On a smaller scale, North Carolina State University recently found that more robust metadata for a collection of photographs increased use of the material fourfold (<https://staff.lib.ncsu.edu/confluence/display/MNC/Evaluating+the+effectiveness+of+manual+metadata+enhancements+for+digital+images>).

To achieve these benefits, federal agencies, publishers, and scholarly and professional societies should provide specific incentives in terms of grant dollars or other methods for those initiatives that explicitly re-use or build on previous work for which open metadata is available, support the creation and/or growth of both institutional and disciplinary repositories, and support the creation and maintenance of registries for the identification of repositories that have open metadata and peer-reviewed content resulting from federally-funded research.

Many early metadata and content sharing efforts, such as the Flickr Commons experiments at Library of Congress, relied on “push” methods that required those responsible for the metadata and content to actively place selected material in individual external systems to enable their discovery and use in those environments. To be sustainable, the sharing of metadata and content beyond institutional and

project borders must in contrast be done via a “pull” model, where all structured metadata and content from a given repository is made available in standard formats via standard technical protocols to any service that wishes to reuse it. This level of openness will allow web-scale escalation of discovery and re-use of federally funded scholarly work. Federal agencies, publishers, and scholarly and professional societies should enact policies and support networks for the open sharing of metadata about peer-reviewed publications resulting from federally funded scientific research via pull methods, to enable this scaling up of metadata sharing. Granting agencies should introduce specific requirements for the creation and dissemination of open metadata. Federal agencies, publishers, granting agencies, and scholarly and professional societies should actively work with and publicly support others’ work with standards bodies to develop and maintain protocols for the open sharing of data.

The metadata and content shared via these methods will support specific actions that will promote the re-use of federally funded research and significantly contribute to the creation of new knowledge. These might include:

- Identification by researchers of studies related to theirs, for purposes of replication, refutation, or enhancement;
- Location of other work done on top of a specific data set of interest;
- Creation of domain-specific repositories of research work, like ArXiv.org;
- tracking of the progression of specific scholarly or scientific ideas over time;
- Integration of geospatial data from federal sources into new research and creative work;
- Improved quantification of the impact of federally funded research

There are a number of mature metadata standards for the description of scholarly papers and research data, for example, the Data Documentation Initiative (DDI) in the social sciences. These standards typically encompass the collection of metadata that support the specific actions named above. These metadata elements include standard bibliographic metadata such as title, author, and subject; funding details; data collection/generation methodology; relationships to published work and other data sets; and domain-specific data relevant to the type of research done.

There are also a number of emerging standards in this area, many of which are aimed at sharing metadata and content on top of core web standards. One primary example of this is experimental work at Johns Hopkins University to use the Open Archives Initiative Object Reuse and Exchange protocol (OAI-ORE) to facilitate the formal publishing of research findings in journals while simultaneously archiving these papers and the raw data generated by the research (<https://wiki.library.jhu.edu/display/DATAPUB/Home>, [http://www.openarchives.org/ore/meetings/hopkins/presentations/Tim\\_DiLauro\\_ORE.pdf](http://www.openarchives.org/ore/meetings/hopkins/presentations/Tim_DiLauro_ORE.pdf)).

#### **#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 public benefit by mandating that all publicly funded research results (using the final, peer-reviewed, edited copy) are available free of charge to the public on federally managed and supported repositories. We have several years of experience using PubMedCentral as such a repository. This experience can guide us to improvements going forward. A few suggestions include:

- Making the deposit of articles easier and clearer, whether it is the author or publisher who does it, to ease the burden of compliance;
- Working with institutions and consortia to facilitate interoperability between article repositories and emerging research management systems like VIVO (<http://www.vivoweb.org> - “an interdisciplinary national network of scientists that will facilitate discovery and collaboration across the country”). These systems foster research collaboration and community engagement based on scientists’ published research output. The University of North Carolina system web portal, REACH-NC (<http://www.reachnc.org>) “enables users to find experts and assets within the state’s higher education and research institutions”. It can be used, for example, to help a local community find a consultant on water quality based on a search of faculty publications and grant awards. The publication part of this portal is obtained from multiple sources including PubMedCentral. Streamlining and expanding this data to include all research publications based on federally funded research would maximize the benefit both to the university and to communities throughout the state and beyond.

**RFI #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 most desirable result is for scholarly publications to be available immediately, with no embargo period. Over 60% of journals already endorse immediate “green” access to the author’s final copy (<http://highwire.stanford.edu/lists/freart.dtl>). Immediate access supports the overall goals of public access by making research results available while they are timely and potentially most useful. For example, at our institution, we have highly visible research programs supporting global health improvements worldwide. A study led by UNC faculty member Dr. Myron Cohen, finding that treatment with anti-retroviral drugs prevents the spread of AIDS, was recently named by *Science* as its “Breakthrough of the Year” <http://www.sciencemag.org/content/334/6063/1628.full>. Such groundbreaking science has widespread relevance in the U.S. and globally. We also have a successful translational research institute that brings our research directly to communities throughout our state and beyond where it can maximize benefits as quickly as possible (<http://tracs.unc.edu>). Such programs receive significant federal research support and their positive impact on society is impeded by embargo periods limiting free access to published research results.

Some faculty have told us that they will publish only in “gold” open access journals because they want their research to be available immediately in developing countries and other places where cost is a barrier to accessing needed information. We have first-hand evidence of critical health information needs that could be addressed by free online access to journal articles in such settings. In a recent encounter, doctors in Uganda needed the latest information about treating complications from kerosene ingestion, a common problem in communities that use paraffin for home cooking. They could not afford to purchase scientific journals to get this information. There are many other countries where, despite widespread poverty, they are not eligible for subsidized journal access through programs like HINARI (WHO). The same needs exist in poor communities throughout the U.S. Open access is not just good for developed countries with large funded research enterprises; it can provide lifesaving assistance around the world and leverage U.S. funded research where it can have maximum impact.

Embargo periods stand in the way of fully realizing the benefits of these and similar programs in which the federal government is heavily invested. If embargo periods are needed to protect publisher profits they should provide data demonstrating this need. Libraries with large research collections will continue to subscribe to the top journals as long as their content is high quality and serves their users' needs. At our institution this is the case despite having suffered major budget cuts in recent years. Embargo periods may be considered important to protect the original work of authors. However, the Creative Commons license (<http://creativecommons.org/>) has gained widespread acceptance. It provides a model for others to reuse and build upon one's original work as long as attribution is given to the original author.

Sincerely,



Sarah C. Michalak  
University Librarian and Associate Provost for University  
Libraries