

Response to Request for Information: "Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research," November 2011

January 8, 2012

Prudence S. Adler

Association of Research Libraries

Washington, D.C.

Summary

Thank you for the opportunity to comment on "Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research." These comments are submitted on behalf of the Association of Research Libraries (ARL). ARL is an Association of 126 research libraries in North America. These libraries directly serve 4.6 million students and faculty and spend \$1.4 billion annually on acquiring information resources, of which 62% is invested in access to electronic resources.

Enhancing public access to federally funded research results is a priority for ARL and its member libraries because such policies are integrally tied to and support the mission of higher education and scholarship. ARL believes that extending and enhancing public access policies to federally funded research to other science and technology agencies will drive scientific discovery and innovation, and promote economic growth. Extending enhanced public access policies to other federal agencies is long overdue.

Question 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?

Comment 1) There are a number of steps that agencies should take to grow existing and new markets relating to access and analysis of peer-reviewed publications resulting from federally funded scientific research. All peer-reviewed articles resulting from publicly funded research should be freely available immediately so that scientists, researchers, students, teachers, citizen scientists, and members of the public can utilize these resources. Importantly, for these uses to be the most effective, accessibility must include the ability to text and data mine, perform computational analysis, and create new derivative works—all executed with no restrictions, for both non-profit and for-profit purposes. It is time to take full advantage of networked, information technologies in order to spur innovation, advance science, and grow new markets.

Despite the growing market share of open access journals, a large percentage of federally funded, peer-reviewed research results are still only available via subscriptions, or sometimes through the purchase of individual articles at a very high

cost. This marketplace model significantly limits access to those who could both conduct research and design new tools and services and, yet, are handicapped by cost and access barriers. There is ample evidence that openly available data and research resources leads to more research, quickens the pace of that research, and yields greater commercialization and development of new tools and services. For example, two reports described below provide clear evidence that openly available resources with no reuse restrictions promoted economic growth and created new jobs and markets. As noted in the Battelle Technology Partnership Practice report, *Economic Impact of the Human Genome Project*, “the \$3.8 billion the U.S. government invested in the Human Genome Project (HGP) from 1988 to 2003 helped drive \$796 billion in economic impact and the generation of \$244 billion in total personal income. In 2010 alone, the human genome sequencing projects and associated genomics research and industry activity directly and indirectly generated \$67 billion in U.S. economic output and supported 310,000 jobs that produced \$20 billion in personal income. The genomics-enabled industry also provided \$3.7 billion in federal taxes during 2010” (http://www.battelle.org/spotlight/5-11-11_genome.aspx).

The link between publicly available research resources, innovation, and commercialization was evident as early as 2002 in a study by Peter Weiss of the National Oceanographic and Atmospheric Administration, “Borders in Cyberspace: Conflicting Public Sector Information Policies and Their Economic Impact.” Three key findings in the report concluded that:

- In Europe, there was little commercial meteorology or weather risk management activity because most European governments did not have open access policies resulting in data being readily, economically, and efficiently available.
- Since the size of the US and EU economies were approximately the same, there was no reason for the European market not to grow to the size of the US with the accompanying revenue generation and job growth.
- A significant contributor to the disparities in weather risk management activity was the difference in information policies between Europe and the United States. In the US, there were no restrictive laws or policies that limited the commercialization of government information.

By making federally funded research results publicly accessible, new audiences and new innovators with differing perspectives are able to benefit from such access. Yet by having much of the federally funded, peer-reviewed literature behind subscription barriers, we are severely constraining our US competitive advantage and our country’s needed investments in STEM education. Extending public access policies that permit full use and reuse rights with no cost barriers will significantly enhance STEM education, level the playing field, and generate more economic growth and job creation in diverse new areas, as seen in the weather risk management and genomic industries. For example, recently we have seen the emergence of new services such as Google Scholar, BioCreAtivE, CoPub, PubGene, and more.

There are deep linkages between openly accessible federally funded, peer-reviewed

research literature and scientific productivity. Research has shown that open access to research literature provides many benefits to science and discovery. For example, it expands the use of research papers, thus increasing citations and the ability to build on the work of others. Previous studies of over a dozen disciplines have shown that open access articles are cited 50–250% more often than those behind subscription barriers (<http://opcit.eprints.org/oacitation-biblio.html>). Reproducibility and building on the work of others are integral to science, and they are also necessities in this new budget environment.

Similarly, an article by Furman and Stern compared citations in follow-on research using materials from Biological Resource Centers (BRC) versus research in closed archives. The authors concluded that articles based on BRC materials received 220% more citations and they were 3–10 times more cost effective in increasing funding of BRCs than funding new research. (<http://www.nber.org/papers/w12523.pdf>). In addition, as described in the paper by Murray, Aghion, Dewatripont, Kolev, and Stern, who evaluated follow-on research done under the auspices of the NIH Public Access Policy, there was “a substantial increase in the rate of exploration of more diverse research paths” (<http://www.nber.org/papers/w14819.pdf>). Another study by Heidi William compared publication and commercial developments resulting from Celera’s intellectual property policies of the human genome and those policies of the US Government. The author concluded that Celera’s intellectual property policies had a negative impact on subsequent research and product development in comparison to the use of Government resources that were in the public domain ([#](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1648013)).

Open access to research resources sparks new approaches to scientific discovery, particularly across scientific disciplines, and such access is especially critical as science is increasingly interdisciplinary and global. As more countries and funders implement open access policies (the United Kingdom being the most recent, “Innovation and Research Strategy for Growth,” 12/2011) the US Government must construct comparable policies for the global scientific enterprise to be as effective as possible in order to address the grand challenges of the 21st century in areas such as health, clean energy, national security, education, and life-long learning.

It is time to reap the benefits of the enormous investments that the US Government has made in cyber and information infrastructure. It is widely understood that these investments are central to advancing science, education, innovation, and our competitive marketplace. And these investments have given rise to new forms of research, allowing scientists to be more productive and explore new research pathways via computational research and analysis. A recent report by the US Food and Drug Administration, “Driving Biomedical Innovation: Initiatives to Improve Products for Patients,” details the many advantages of effectively utilizing computational systems and tools to drive scientific research, innovation, and commercialization.

“The ability to integrate large data sets across multiple clinical trials, post-market surveillance data, and pre-clinical data will enable FDA to generate new insights into a variety of important issues confronting medical product development and use. Examples of such insights include the identification of patient subsets who do or do not respond to a specific therapy during a clinical trial, which has the potential to drive

personalized medicine; identification of patient subsets with differential safety profiles, efficacy, or side effects related to age or gender; evaluations of standard of care; analyses of disease progression; assessment of current endpoints based on aggregated data; and potential to generate better endpoints and insight into placebo effects. This work, which will address broader scientific issues, is intended to impact whole product classes and therapeutic areas and will be central to driving innovations in medical product development and basic research”
<http://www.fda.gov/AboutFDA/ReportsManualsForms/Reports/ucm274333.htm>).

Research has shown that if the US Government were to adopt an open access policy, it would result in a five-fold increase in the return on investment. Given the current and anticipated budgetary environment, it is difficult to understand why the US Government would not adopt an open access policy. The net gain of extending an National Institutes of Health (NIH)-like policy to other agencies is estimated to be \$1.5 billion (www.arl.org/sparc/publications/papers/vuFRPAA/index.shtml).

There has already been investment in needed infrastructure by the NIH. Extending an NIH-like Public Access Policy to other federal agencies could be accomplished in a cost-effective manner by building on NIH’s investment in PubMed Central. Such an approach would avoid duplication of effort and is the most logical given the current budgetary environment. For example, the annual cost of providing access to the results of NIH funded research is between \$3.5–\$4.6 million dollars. For the nominal cost of one one-hundredth of one percent (.0001%) of NIH’s overall budget, more than 500,000 users per day from public and private domains have access to a database of over 2 million articles.

One important driver of the NIH Public Access Policy is accountability with regards to NIH’s research portfolio. Maintaining a repository of all NIH-funded research results provides the agency with information and analyses concerning the investments it has made in biomedical research. The NIH Public Access Policy supports science-based budget determinations and assists NIH, Congress, and the biomedical research community in understanding the outcomes of the funded research and how best to identify and target new areas of research to support.

In order to maximize the investments in cyber and information infrastructure, advance science, and promote innovation, free immediate access with full reuse rights to federally funded research literature would achieve the most benefits. There should be no restrictions placed on use of this literature or on who is able to use these federally funded information resources. This would be consistent with existing federal policy, the Paperwork Reduction Act and Circular A-130, concerning government information. If an embargo period is deemed necessary, it should be as short as possible.

Question 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?

Comment 2) Key to the success of advancing research, and spurring innovation and commercialization, will be to provide unfettered access to federally funded research resources and permit the widest possible use within the law. This is possible by utilizing Creative Commons CC-BY or comparable open licenses that work within copyright law and are already widely employed by individuals in all sectors. Use of these licenses permits the user full use rights to mine data and text, and manipulate, reuse, and integrate data and information in publicly accessible digital repositories. The use of CC-BY licenses or open licenses should be integral to a new federal open/public access policy.

As the White House considers a new federal open/public access policy, it is essential that the results of federally funded research be accessible in the most effective manner. So for example, if an embargo is deemed necessary, it should be as short as possible. And once the embargo is lifted, then full reuse rights should be associated with the research literature. Such an approach takes into account the needs and interests of all stakeholders. Regardless of where the publications reside, full reuse rights are essential elements of an effective policy.

Question 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?

Comment 3) The US Government has a long history of ensuring that there is long-term preservation of and access to works via centralized deposit. For example, through a provision in the Copyright Act, printed copyrighted and public domain works are placed on deposit at the Library of Congress. Beginning in 2010, the Library extended this deposit requirement to include electronic-only serials. The National Library of Medicine has been providing long-term preservation of and access to biomedical information for 175 years. More recently, NIH implemented the NIH Public Access Policy, which is a natural continuation of this role. It is appropriate and necessary for the US Government to ensure that the long-term preservation of and access to these resources is undertaken and with appropriate use rights for the Government and users alike.

As more and more institutions and organizations establish digital repositories, there will be many sites providing access to federally funded research literature, nationally and internationally. For example, PubMed Central is one of many sources for the biomedical literature it archives once any embargo period for an article has expired. Any US policy must ensure that these repositories of federally funded research resources are interoperable and accessible with appropriate use rights both now and in the future, regardless of who is curating these resources. As we have learned, long-term preservation of and access to digital resources requires use; dark archives are not an option. To ensure that there is not deterioration of these digital resources and that there is a valid record going forward, continuous use is required.

Innovative public/private partnerships may emerge that will allow for the creation of new tools and services built upon these federally funded research resources. And as these partnerships emerge, clearly delineating roles and responsibilities will be key. Importantly, it will be critical to stipulate that if a provider for some reason is unable to meet its obligations of service—either short-term or long-term—a migration path should be in place to recover the resources. This latter point is especially important given the recent study by Cornell University and Columbia University that found that the majority of their journal holdings are not archived by LOCKSS and Portico (<http://2cul.org/node/22>).

Question 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?

Comment 4) Libraries and many universities have a long history of partnering with others to ensure the long-term preservation of and access to research resources. For example, the Inter-University Consortium for Political and Social Research (ICPSR) is comprised of about 700 academic institutions and research organizations. “ICPSR provides leadership and training in data access, curation, and methods of analysis for the social science research community. ICPSR maintains a data archive of more than 500,000 files of research in the social sciences. It hosts 16 specialized collections of data in education, aging, criminal justice, substance abuse, terrorism, and other fields” (<http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>).

Another example is ArXiv, hosted by Cornell University Library. It is an archive of 726,955 electronic preprints of research papers in the fields of mathematics, physics, computer science, quantitative biology, statistics, and quantitative finance (<http://arxiv.org/>). More recently, HathiTrust Digital Library was established and is a partnership of major national and international research institutions and libraries working to ensure that the cultural record is preserved and accessible in the future. There are more than 60 partners in HathiTrust (<http://www.hathitrust.org/>).

These partnerships demonstrate the commitment of research libraries and universities to the long-term preservation of and access to cultural and scientific records. Key to their success includes requiring the appropriate terms and conditions for long-term preservation, curation, interoperability, and use rights. In addition, these partnerships show that universities and libraries have expended and will continue to expend a significant amount of resources—staff expertise, financial support, and infrastructure investments—to ensure that these resources are publicly accessible in an effective manner both today and in the future.

Question 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?

Comment 5) Well-documented metadata is an important means to enable use, reuse, and analysis of the research literature and data. All of these uses should be machine-readable and interoperable. Readers, both human and machine, must know the terms and conditions and provenance under which this research may be used. Thus federal agencies should understand the important linkages between metadata and achieving a robust open/public access policy for science and technology-related agencies.

Given the extensive community efforts already underway, there is deep value in building upon existing standards such as Dublin Core, OAI-PMH, DataCite Metadata Schema, and Europeana Sematic Elements. In addition, efforts such as ORCID provide important contributions to this arena. ORCID seeks to resolve “the author/contributor name ambiguity problem in scholarly communications through the creation of a central registry of unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID and other current author ID schemes. These identifiers, and the relationships among them, can be linked to the researcher's output to enhance the scientific discovery process and to improve the efficiency of research funding and collaboration within the research community” (<http://www.orcid.org/>). Finally, there is value in looking to other existing organization such as the National Information Standards Organization, a non-profit organization devoted to collaborative standards development amongst content publishers, libraries, and software developers.

Question 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?

Comment 6) Ensuring that all federally funded research results are accessible, and available in an effective and timely manner, will maximize the benefits to the scientific enterprise and to the public. For any open/public access policy to be successful, there must be consistency of requirements and mandates. It will be difficult for research universities to comply with multiple and differing mandates, in part, because a federal open/public access policy may involve multiple research funding agencies. Research universities have faculty members and researchers who hold grants from all or several federal funding agencies, and some of them have grants from multiple agencies concurrently. To the extent practicable, uniform requirements and procedures regarding deposit of peer-reviewed literature should be established across all funding agencies, as uniformity of deposit requirements will reduce the complexity and cost, while at the same time increase the rate of compliance. Ensuring relative consistency across agency policies is one key element to ensure a valuable return on investment and foster a culture where sharing of these resources continues to promote the interests of science.

To that end, open/public access policies should build upon existing policies and protocols for deposit of peer-reviewed literature, should promote development of new tools and services, and should integrate federally funded research grants and resources into the grants management systems within the agencies and in the research institutions. Such measures build on accountability metrics that many research universities are actively integrating into the research enterprise. These metrics assist the research university in detailing their research outputs and the local, state, national and international value of their institution. It is in this context that many institutions have invested in digital repositories so that the research results of their institution are publicly available, and for their community of users to build upon these repository resources as teaching tools and to advance scientific discovery.

Question 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?

Comment 7) There are other important types of scholarly communications beyond the peer-reviewed research literature. Monographs and book chapters, conference presentations, theses and dissertations, working papers, and datasets are also increasingly being made available via open access or public access policies. Policies covering ETDs (electronic theses and dissertations) are also common, well developed, and generally supported by students as well as their faculty advisors. Since ETDs are authored by students rather than faculty, ETD policies are usually developed through a different process than policies targeted at faculty research outputs. Since there are different terms and conditions associated with each of these educational materials, it will be important to distinguish the various approaches to each type of scholarly output.

The related RFI concerning data policies indicates that data policies may be differentiated from peer-reviewed literature and other types of scholarly output as different terms and conditions may apply. Nevertheless, data is central to the scholarly and research enterprise and should be treated equally in terms of importance to the scholarly record and tenure and promotion.

Question 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?

Comment 8) Isaac Newton's statement that he "stood on the shoulders of giants" aptly describes how advances in science build on prior knowledge and the sharing of information. It is time to accelerate such advances by significantly decreasing or eliminating embargoes to currently available, published research resources. Nationally and internationally, embargo periods of 12 months or less are the standard for journal

publishing (<http://highwire.stanford.edu>). If it is necessary to accommodate those journal publishers whose marketplace models depend upon subscription revenue, the US Government should adopt a policy with an author embargo period that is as short as economically feasible, but no more than 12 months. It is important to note that the NIH Public Access Policy (with an embargo period of 12 months) is not representative of international biomedical funder policies. A six-month embargo is now standard (<http://roarmap.eprints.org/>).

Any determination of the need for a different embargo period must be based on data provided by a subscription-based publisher that shows a negative market impact resulting from the open/public access policy. In so doing, a range of factors should be considered. First, the pricing history of the journal and other journals within that discipline must be compared. Second, the impact of subscriptions via bundles vs. single journals in a discipline should be considered. Third, peer-reviewed journals include information well beyond articles stemming from federally funded research. They include articles based on other funding sources and also include information about conferences, professional development, and more. As a result, it will be important to identify the percentage of articles based on federally funded research in a subscription-based journal to truly understand the need for a different embargo period. Fourth, it is incumbent upon a subscription-based publisher to provide data on the revenue that results from long-tail citation articles.

Finally, the economy has significantly affected research universities, and as a result has impacted research library budgets. This is particularly true for public institutions as state budgets face weak economic growth, receive fewer federal dollars, and local governments are unable to keep pace with demands for services. This all translates into fewer and fewer dollars from states to their public institutions. And more reductions are anticipated. ARL conducted a survey of its members between 2008–2010 to better understand the fiscal environment. Overall, over 79% of ARL member libraries had flat or reduced annual budgets from FY 2008–09 to 2009–10. Of the 61% that had real dollar budget reductions, the maximum budget cut was a striking 22%. ARL libraries continued to face budget reductions in 2011 and planned for permanent reductions in both staff and collections resources.

Understanding the relationship between these fiscal challenges, indeed all of the factors noted above, and subscription cancellations is very important. All of these factors and in particular, the “new norm” of budget realities factor in to how research libraries approach collection development. If a different embargo period is considered due to a *perceived* negative marketplace impact, all of these factors must be considered to understand the real impact of the embargo period.