

## **Response to Request for Information Public Access to Peer Reviewed Scholarly Publications Resulting from Federally Funded Research**

Richard E. Luce, Vice Provost and Director of University Libraries  
Emory University  
Atlanta, Georgia  
[REDACTED]

### **Comment 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 the U.S. economic growth and improve the productivity of the American scientific enterprise?]*

Because of the benefits both to scientific productivity and the economy, federal agencies should adopt policies to make articles resulting from federally funded research publicly accessible. Scientific research creates marketplace demands for labor, products, services, and further research. Private investment and commercial applications that result from cutting-edge research will spur economic growth, potentially creating jobs across broad sectors of the economy. With public access to previous articles, scientists can avoid repeating research, and can discover new avenues of inquiry. Finally, public access means that taxpayers have access to the results of research their dollars funded.

Science is, by its very nature, cumulative—all new knowledge builds on earlier findings. The ability to access and use previous research accelerates scientific discovery and innovation. Open access to research articles increases citations and follow-on research, diversifies pathways in research, and accelerates applications in industry and commercial settings. It is critical that works be free for access and re-use (with minimal embargo restrictions), to enable a full range of activities including text and data mining, computing, and the creation of derivative works. Enabling not only open access but also full reuse means additional value—experiments are not repeated, and research not re-done. This, essentially, enables us to “do more with less.”

Making research articles open access also enables the inclusion of machines as a new category of reader, opening up research pathways and making new connections possible. Scientists can not only work faster, they can work smarter, deploying new search, discovery and semantic tools. For example, the National Center for Biotechnology Information (NCBI)'s PubMed Central (PMC) provides research articles in a machine-readable format. This means not only that these articles are openly accessible, but also that articles in PMC can be a gateway to some 40 databases and 13 trillion bytes of NCBI data.

Such access can advance scientific discovery immeasurably. (<http://www.hhs.gov/asl/testify/2010/07/t20100729c.html>).

Because of existing infrastructure, the relative costs of public access policies are manageable. The example of PubMed Central is instructive. After start-up costs of \$500,000, annual operating costs (including ingest, refinement of submission system and search tools, and national office and help desk staffing) run to \$3.5 - \$4.0 million per year. This is only a fraction of NIH's \$30 billion/year overall budget—about 1/100<sup>th</sup> of 1 percent. The benefits of an open access policy like the NIH policy are estimated to encompass approximately 8 times the costs. The net present value gains of expanding an NIH-style policy to all other U.S. science agencies is estimated to be on order of \$1.5 billion—with 60% of that accruing directly to the U.S. economy. Open access policies for other agencies can be implemented in a cost-effective manner by leveraging the PMC infrastructure that exists for making articles open access. Supplementing existing access points with additional manuscripts would represent only small incremental costs.

Ideally, the type of access enabled by such policies should be immediate and full—with embargoes either eliminated or minimized. Since investment in science should create jobs now—not twelve months from now—restrictions should be minimal.

Scientific progress leads to economic development, and limiting this progress by granting access only to a privileged few hinders it. Scientific research drives new discoveries and technologies, which affect the economy by creating new business needs, services, and jobs. When considering the costs of doing duplicate research and pursuing blind alleys, the potential cost saving in dollars and researcher hours would likely run high. By saving time and money, more research could be done for the same expenditure.

## **Comment 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?]*

Adopting public access policies does not mean turning copyright law on its head. Agencies can work within the current copyright framework to provide public access to scientific research. Both scientists and the public, however, deserve access to and full use of these articles sooner than the current term of copyright allows. Access can be accomplished, while still protecting the intellectual property interests of all stakeholders, by implementing appropriate licenses (e.g. Creative Commons CC-BY licenses). Such licenses enforce, rather than subvert, current copyright law.

Policies resulting in “read-only” databases should be avoided. Mechanisms to enable full use of research—including distribution, reuse, text mining, computation, etc.—should be included in any government public access policy. In attempting to balance stakeholder

interests, a useful strategy to consider would be a tiered approach. First, articles could be provided for a period where current rights apply; after a period of time, articles could be accessible with full reuse rights under an appropriate open license (CC-BY or similar).

### **Comment 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?]*

The federal government is in a unique position to ensure that publicly funded articles are preserved and made accessible. Any public access policies that are developed must give the federal government adequate rights to archive and distribute publicly funded articles. PubMed Central illustrates the benefits of both a centralized repository and a government mandate to open access. In its first few years of existence, contribution was voluntary. However, only 5% of eligible articles were submitted by their authors. While journals also submitted a portion of articles, under this voluntary system, only 19% of articles eligible under the policy were included in PubMed Central. Now that submission is mandatory, NIH is able to increase content while refining their information systems for ease of use, navigation, and information finding.

A federal public access policy could involve multiple repositories, but these repositories must be able to ensure long-term preservation, access and usability. Interoperability and long-term preservation could be maintained by third parties, encouraging partnerships between the public and private sectors. Simply providing the government with a copy to put in a “dark archive” is not a viable solution; without regular use, archival veracity cannot be ensured. Library experiences have shown that regular access and use of digital materials is a crucial element in effective long-term preservation.

The drawback of a decentralized approach is the potential for failure. Current attempts at archiving scientific literature are inadequate. In a widely cited study, Cornell and Columbia report that only about 15% of their combined journal holdings are currently archived by the LOCKSS (Lots of Copies Keep Stuff Safe) initiative and Portico archive combined (<http://2cul.org/node/22>). Despite this dismaying figure, the Cornell/Columbia efforts are considered to be a relatively successful digital preservation initiative. LOCKSS and Portico have some degree of overlap, but both services preserve titles uniquely. There is much room for improvement, which could arise from federal stewardship of research products.

### **Comment 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 could be a vital component of a federal public access policy, provided repositories meet conditions for public accessibility, use rights, interoperability, and long-term preservation. Since no single site should be the sole point of access for research articles, partnerships between federal agencies and private entities could be the key to ensuring a viable system. Universities and libraries have extensive experience and existing archive infrastructure, and should be actively encouraged to partner with federal agencies.

One model for public-private partnership is PubMed Central. When PubMed Central was launched in 2000, a number of publishers collaborated, depositing their journals in the database. Using these submissions, NLM developed the structured digital format for representing digital journal articles—the NLM DTD. In turn, this format is being adopted by publishers and libraries (and is in the process of becoming a NISO standard).

There are also good examples of partnerships among academic research institutions, and companies, providing access and preservation. For example, the DRIVER initiative is a European network of digital repositories housing 3,500,000 scientific publications—a consortium of sorts of 295 open access repositories, from 38 countries (<http://www.driver-repository.eu/>). HathiTrust is a partnership of leading research institutions and libraries, with sixty partners worldwide (<http://www.hathitrust.org/>). Both initiatives seek to preserve and provide access to content, with both a central location and integration into partner libraries' local systems.

#### **Comment 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 for 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?]*

Metadata should facilitate use, reuse and analysis of published works, and must be machine-readable and machine interoperable. Metadata in its current form is a means for enabling actions, rather than simply embodying description of items. Existing standards (Dublin Core, OAI-PMH, Datacite Metadata Schema, Europeana Semantic Elements, ORCID, I-2, Counter/SUSHI) may help inform metadata requirements, and agencies currently working to improve metadata interoperability (e.g. NISO, LoC) will be essential.

Furthermore, one critical element of metadata is that it must be coupled with an API for standards-based data exchange. It is not enough to create metadata—descriptors must be easily shareable among institutions and repositories. The importance of shareable metadata indicates that libraries must be vital partners in this process.

#### **Comment 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 of stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?]*

In order to maximize the benefit of public access policies to taxpayers, while minimizing burdens and costs to stakeholders, uniform requirements and procedures surrounding the deposit of articles should be established across all funding agencies. Institutions and researchers often hold grants from multiple agencies concurrently, and common requirements will reduce complexity and increase compliance. Policies should take advantage of existing protocols (for example, the SWORD standard, <http://swordapp.org/>) to facilitate automatic deposit of manuscripts to multiple repositories. Agencies can integrate articles with grants management systems, improving agency accountability and compliance while providing increased information to the public on research results.

The PubMed Central database currently contains over 2 million articles, and is used by more than 500,000 users per day, underscoring a deep demand for this information in the public sector. NLM statistics indicate that while 25% of users of PMC are from universities, 40% are private citizens or those using personal internet accounts, and 17% are from companies (the remainder being government or other users). PubMed Central is a broad-based repository used not only by researchers and students, but also clinicians, patients and their families, and entrepreneurs. Other federal agencies can take these statistics to indicate that the public values access to scientific research.

A major benefit of public access policies is that taxpayers can access the results of research their tax dollars fund—immediately and fully. Simply put, opening access to publicly funded research articles increases the public's return on their investment in that research. Since, by paying taxes, Americans are the investors in federally funded scientific research, public access policies maximize the value of investment in research.

### **Comment 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?]*

Along with journal articles, a host of other materials are good candidates for public access. For example, educational materials (such as book chapters, texts, and conference proceedings) could be made readily accessible to the public. Making these materials open access will improve the quality of education, allowing cutting-edge research to be quickly incorporated into the teaching process. Additionally, much as in the case of journal articles, more access will mean better science. However, the policies under which these materials are made accessible will likely differ from those directed at journal articles. While authors are not paid for journal articles, they may make significant royalties from textbook

chapters, for example. Since scientists and scholars may rely on this income, policies will need to reflect these differences in the current scholarly publishing apparatus.

### **Comment 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?]*

To optimize both scientific progress and economic growth, immediate access is best. However, journals do rely on subscription income. An embargo period of up to 12 months has proven to be an effective accommodation of journals' economic models, while still resulting in improved access to scientific research. It is important to note that no publisher has provided empirical data indicating that this embargo system (currently used by NIH and numerous funders worldwide) harms journal revenues or drastically alters their economic models. Embargoes of 12 months or less have been adopted by hundreds of journals (<http://highwire.stanford.edu/lists/freeart.dtl>) and research funder policies around the globe (<http://roarmap.eprints.org/>).

Public access to scientific research does not need to interfere with journal revenues. Even publishers who initially balked at opening access to back content have changed practices. In the name of scholarly exchange of knowledge, the Royal Society recently opened its entire historic journal archive—including the first ever peer-reviewed journal—dating back to 1665, adopting a 12-month embargo period. (<http://royalsociety.org/news/Royal-Society-journal-archive-made-permanently-free-to-access/>). They noted that this back file accounted for less than one half of one percent of their overall publishing revenue.

The claim that public access policies negatively affect publishers (presumably because of cancellations) often does not take into account other factors, including growth of directly competing journals, price and pricing history, revenue resulting from "long-tail" citations, impact of publisher-required bundles versus single journals, and library budget numbers. These and other market conditions affect journal cancellations, and must be accounted for in any analysis of lost revenue.

In fact, studies have shown that open access articles actually have the advantage in both visibility and citations. Articles whose authors supplemented subscription-based access by self-archiving their final drafts gained citations, in some cases doubling citation rates for articles that are not open access. Thus, if taking into account both public benefits and scientists' interests in their own research, open access articles have an advantage.