



December 8, 2011

White House Office of Science and Technology Policy
Request for Information on public access to peer-reviewed scholarly publications
resulting from federally-funded research.

RESPONSE from the Duke University Libraries.

(1) Are there steps that agencies could take to grow existing and new markets related to the access and analysis of peer-reviewed publications that result from federally funded scientific research? How can policies for archiving publications and making them publicly accessible be used to grow the economy and improve the productivity of the scientific enterprise?

What are the relative costs and benefits of such policies? What type of access to these publications is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise?

The most important and effective policy to accomplish these goals would be to require immediate public access to research publications that stem from federally-funded research via an easily searchable database or set of federated databases. Such public access should be free of charge to all users and should include a broad range of reuse rights so that users can build on and innovate from the research that they find. This greater access to the huge amount of scientific research that the federal government funds would result in faster follow-on research, new products and services and more jobs. It is a relatively simple and cost-effective way to dramatically increase economic growth and global competitiveness.

In his Directive on Open Government, President Obama stated that government should be transparent, participatory and collaborative. Increasing public access to federally funded research will serve all of these goals. It will also make government more accountable for how it spends taxpayers' money; such accountability is also a benefit the President promised as part of the open government initiative.

In order to achieve these benefits, a mandate for public access should be more comprehensive and include a broader set of reuse rights than the National Institutes of Health public access policy currently does. Commercial reuse, for example, is an important component in using public access to foster commercial innovation and growth, especially because it can benefit smaller and start-up companies. Reuse by computers -- the ability to treat large repositories of text as data from which to discover new connections, collaborations and paths for exploration -- is another vitally important advantage of broader reuse rights.

A broad and comprehensive collection of federally-funded research publications across a wide variety of disciplines will foster discovery and thus support innovation. The

ability for more people to discover unexpected opportunities and interdisciplinary connections is a key to this benefit. Public access will foster subsequent research at a much faster pace than is currently the case. Also, public access takes account of the near certainty that there are unexpected readers and users who can innovate in ways not anticipated by the original researchers or their funders. These unanticipated users, who include commercial innovators, small companies, under-resourced researchers and potential collaborators, offer the best hope for true innovation. Nearly every experiment in open access exposes these opportunities, and federal agencies should design public access mandates to exploit them.

One research study that confirms the benefits of public access is the Working Paper by Heidi Williams of the MIT Economics Department called "Intellectual Property Rights and Innovation" in which she studied the follow-on R&D based on two types of gene-sequencing research, one proprietary and one publicly accessible. Williams found that the proprietary efforts suffered a 30% reduction in subsequent scientific research and product development.¹ This low-growth approach is now the norm for most federal agencies, and that must change.

Another study, by the Joint Information Systems Committee in the UK, found that the cost/benefit ratio for a static-state model of public access to funded research through national repositories was 122, or a five-fold return on investment.² Other countries are rapidly discovering the advantages of public access to scientific research and information, and the U.S. must do so as well in order to compete and grow.

These benefits can be obtained at a very reasonable cost. The National Institute of Health reports that the cost of making its funded research available to the public is approximately 4 million dollars out of a 30 billion dollar budget. Their database is used by over a half million users each day, suggesting that there is great demand for this information and that a very high return on investment can be achieved at modest cost. As access and reuse is increased beyond the NIH model, that return on investment will rise even higher.

(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?

When considering how best to manage the intellectual property rights associated with federally-funded research, it is important to remember that the initial rights holders,

¹ Professor Williams' paper, in the National Bureau for Economic Research working papers series, is found at <http://www.nber.org/papers/w16213.pdf>. (accessed Nov. 23, 2011).

² The JISC report can be accessed at <http://www.jisc.ac.uk/publications/reports/2009/economicpublishingmodelsfinalreport.aspx>.

for whom the system is structured, are the researchers and authors themselves. Attention to copyright, for example, is important insofar as it serves as an incentive for new discovery and authorship, but it must be recognized that copyright can also stand in the way of such innovation. A balanced approach that is focused on the incentives for new creation and discovery is called for.

In the academic world especially, the incentive purpose of copyright is short-circuited by the fact that authors are not paid when they publish their works. Instead, the rewards for doing research and publishing scholarship are all provided by the universities and the funders operating entirely outside of the structure of copyright. The benefits from the copyright monopoly thus are absorbed entirely by publishers who have a limited role in incentivizing new science. Whatever limited concessions must be made to publishing interests, therefore, should be weighed against the need for rapid and diffuse advancement of science, and should not be allowed to interfere with that goal.

The copyrighted works that would be the subject of public access policies benefit their authors most when they are widely available. For most scholars, copyright is a concern only insofar as it serves their desire to have the greatest possible impact on their research field.

For this reason, public access does not threaten the incentives for the creation of scholarship at all. In fact, it enhances them because it supports faster research, more innovation and greater impact on the field. Numerous studies have shown, for example, that open accessibility of scholarly articles increases the number and the pace of citations.³ Many scholars have recognized the benefits of leveraging their copyright ownership to provide greater access and fuller reuse rights than the copyright law itself permits (under fair use), often by using Creative Commons licenses on their works when they are able to do so.⁴ Thus the evidence suggests that the rewards system that actually provides the incentives for a large group of scientific researchers is supported by public access, not threatened by it.

(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 most important feature of a public access policy is the ability to search across a broad and interdisciplinary set of databases. These databases must be

³ The classic study identifying this citation advantage is **Eysenbach, G.** 2006. Citation advantage of open access articles. *PLoS Biology*, 4(5), 692-698. Available at <http://www.plosbiology.org/article/info:doi/10.1371/journal.pbio.0040157>. (accessed Nov. 29, 2011).

⁴ <http://creativecommons.org/licenses/by/3.0/>.

crawled by search engines, since even highly technical scientific research these days often begins with Google Scholar, especially when a new field or sub-field is being broached. Under these conditions, multiple databases are fine. But it is important not to allow private entities to have too much control over access and functionality; doing so risks undermining the very benefits a policy is indeed to support.

In order to support long-term preservation and to provide the necessary level of accountability that is one of the hallmarks of a public access policy, at the very least the government should maintain a mirrored and accessible version of the databases, regardless of who the primary managers of the repositories are. And, of course, the government will be responsible for setting and maintaining the standards under which multiple repositories would operate. Again, government openness and accountability are a key part of public access policies, so ensuring interoperability, access, persistence (commercial firms do go out of business) and uniform standards are key.

In truth, a set of centralized databases may prove both more effective and more economical. The NIH has discovered that running a centralized database for funded articles is not very expensive and yields very high returns, as will be discussed more fully below. Also, a centralized database (whether maintained by the agencies or through partnerships with private entities) makes standardization easier, both for submission and for discovery and use. Authors or publishers would only need to learn (and support services like libraries would only need to support) a single or small set of submission workflows and APIs. And any number of third party services (free or commercial) could develop discovery and use interfaces against a single API, stimulating innovation in the discovery and use sphere.

Where agencies will create and manage repositories, it would also be preferable to use open source software. The ability to innovate the platform for searching and sharing such data may be critically important to enabling effective reuse of these contents. Alternatively, the ability to download the contents easily and to export to another platform might enable such innovation. If there were a way to prevent those who might develop private databases that incorporate this public information from laying hurdles, such as software patents, in the way of others who might engage in such innovation, that would be welcomed.⁵

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?

The models for public-private partnerships that are most relevant for supporting public access to federally-funded research are those with universities. It is an unfortunate fact that, in the area of scholarly communications, commercial publishers have often proved to be unreliable partners because the overall need for profitability has led to resistance to the very features that are key to a successful public access policy.

⁵ This point was made to us by Dr. Anthony So, Professor and Director of the Program on Global Health and Technology Access at Duke University.

Researchers themselves frequently express distrust of the publishers on whom they still must rely for dissemination of their work; it would be unfortunate for a public access policy to begin its life encumbered by this distrust. Universities, both public and private, have simply proven themselves to be better at cooperation, especially over long-term projects. The example of the ArXiv database, which was begun at the Los Alamos National Laboratory and is now hosted at Cornell University, is an excellent example of a working partnership between governmental and private agencies that has already provided a huge benefit to the fields of physics, mathematics, computer science and quantitative biology.

Regardless of where partners are found, standards for access, preservation and interoperability must be maintained. If privately managed repositories can be federated in a way that supports innovation, then such partnerships are fine. But federation does not merely mean the ability to search multiple platforms in a passive way. To really foster innovation, it is important that creativity and experimentation be allowed even as to the tools that will be used in searching the publicly accessible articles. The ability to create new kinds of search tools, and, critically, to compute on the contents of these repositories (text-mining, etc.), must be afforded to users in order to exploit the full range of opportunities for innovative research and economic growth.

(5) What steps can be taken by Federal agencies, publishers, and/or scholarly and professional societies to encourage interoperable search, discovery, and analysis capacity across disciplines and archives? What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities? How should Federal agencies make certain that such minimum core metadata associated with peer-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to Federal science funding?

It is important to understand that metadata is a key aspect of achieving the benefits of open access, since good, rich metadata facilitates the cross-disciplinary discovery and even the ability for machine-reading research that will improve scientific productivity. Metadata based on the Dublin Core should be a baseline standard, and it must be coupled with a linked-data API that supports standards-based data exchange. Again, developing a metadata scheme that can support machine-readability and interoperability is vital. Agencies that have long experience in the development of metadata schema, including NISO and the Library of Congress, should be involved in the creation of appropriate standards. In general the best approach is to look to models that are already successful. Libraries and other groups already have working metadata schemes, and using such a model means that researchers and librarians (as well as automated systems) will already have developed search techniques that work with the scheme. Thus the benefits sought by requiring public access can be accelerated or, properly-speaking, not delayed unnecessarily.

One important function of good metadata standards is the ability to track and analyze usage. The COUNTER project (Counting Online Usage of Networked Electronic Resources) is an important initiative to provide this kind of analysis and

should be a significant partner in any repository development.⁶ This kind of analysis will support accountability and allow federal agencies to quantify the return on investment benefits that will certainly follow from public access policies.

Another way in which metadata description can support the benefits sought through public access is by linking research publications with the data that underlies them. Again, this must be done in a way that is meaningful to readers but also navigable by computers. Unique identifiers and vocabularies that support semantic relationships will make these repositories much more valuable and, again, able to return the kind of benefits that science and commercial innovation depend upon.

(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?

A 2006 report by Professors John Houghton and Peter Sheehan of Victoria University in Australia describes and models the dramatic return on investment that public access offers, showing impact ranges from 25 to 75% on research and development when access to funded research is increased.⁷ The simple fact is that public access is a very good use of tax money, offering an extremely high return on investment, as well as an increase in government accountability. The example of the National Institutes of Health shows that even a public-access policy that relies on an agency-managed repository can be done for a very modest cost, and the benefits of that investment have uniformly proved to be immense. Houghton and Sheehan demonstrated that increasing access and reuse beyond what is possible from the NIH will further increase these benefits.

The most important step that can be taken to minimize cost and maximize compliance is to develop consistent policies. Across agencies and across platforms, the requirements and procedures should be as consistent as it is possible to make them. Consistency will reduce the manpower necessary to educate users and to troubleshoot compliance problems.

Another possible step to take – which would be difficult to maintain at scale but could serve as an initial step – is to consider the option of requiring deposit in any one of an identified set of vetted and trusted repositories, including, perhaps, disciplinary and university repositories, and then harvesting across these repositories in order to create access to the body of funded research. In this model consistency is also paramount, not only in the requirements and procedures for compliance but also in the identification of those repositories that meet the necessary standards.

⁶ <http://www.projectcounter.org/about.html>.

⁷ John Houghton and Peter Sheehan, “The Economic Impact of Enhanced Access to Research Findings,” CSES Working Paper no. 23, Centre for Strategic Economic Studies, July 2006. Available at <http://www.cses.com/documents/wp23.pdf> (accessed Nov. 30. 2011).

(7) Besides scholarly journal articles, should other types of peer-reviewed publications resulting from federally funded research, such as book chapters and conference proceedings, be covered by these public access policies?

The core of a public access policy should clearly be peer-reviewed journal articles, since this is a medium for which there are clear standards and expectations, whose use is familiar to most people, and for which the intellectual property rights situation is well-understood. As noted above, authors of these papers do not depend on royalty payments for an incentive, so participation by the funded researchers will be less problematic in this area than those in which commercial expectations are paramount. Other types of materials will have different conditions, and the IP rights can get very complicated in some cases. Educational objects -- digital objects created to teach a particular concept or point -- would be another type of material to consider, but the standards of peer-review are not yet clearly in place for those works, and the IP rights can be very unclear. So while these educational objects, if funded by federal research money, might well be the subject of a separate mandate crafted for the specific conditions, it is clear that journal articles should be the starting point. The complexity of other materials should not be allowed to delay the implementation of a mandate for public access to funded articles across federal agencies.

(8) What is the appropriate embargo period after publication before the public is granted free access to the full content of peer-reviewed scholarly publications resulting from federally funded research? Please describe the empirical basis for the recommended embargo period. Analyses that weigh public and private benefits and account for external market factors, such as competition, price changes, library budgets, and other factors, will be particularly useful. Are there evidence-based arguments that can be made that the delay period should be different for specific disciplines or types of publications?

It is vital to understand that any embargo necessarily involves a tradeoff that diminishes the possibilities for innovation and growth that a mandate is trying to realize. In economic terms, each day of an embargo period represents lost opportunity costs that decrease the return on investment that is possible for the policy. For this reason, an embargo period is always a compromise, and they should be as short as possible.

Because the conditions and traditions of individual disciplines vary a good deal, it is probably best to allow authors themselves to determine appropriate embargo periods, or to permit immediate public access. In any case, this option should be preserved only with defined limits. For example, in order to avoid excessive negative impact on both the research community and the patient community, embargoes in the biomedical sciences should probably never exceed six months. We note that there is a strong sense among those who study global health issues that the current 12 month embargo allowed by the National Institutes of Health is too long, as indicated in a report funded by the Institutes of Medicine of the National Academies on “The U.S. Commitment to Global Health.”

Several prominent medical research funders have made open access a condition of grant support. The European Research Council, a funding body set up by the European Union to promote research in the region, has also put forward an open access policy requiring its grantees to post all publications to a research repository within 6 months of publication. This marked the first EU-wide open access policy and ERC has stated that it has interest in shortening the 6 month window period in the future. The Wellcome Trust requires submission of scientific publications resulting from its grants into UK PubMed Central within six months of the publication date and even provides funding for the upfront fees associated with publishing in such outlets. Grantees of the Howard Hughes Medical Institute also face a similar requirement to deposit publications in PubMed within six months of the publication date. By contrast, NIH's Public Access Policy remains at twelve months, twice the embargo period accepted by other leading funding agencies.⁸

It is worth noting that many journals, in a variety of disciplines, have adopted embargoes of less than a year, as publishers discover that the loss of revenues that they feared have simply not materialized.⁹

As far as empirical arguments are concerned, we should begin with the recognition that the evidence supporting the value of public access is plentiful and growing all the time. The burden of proof for the need for embargoes should therefore rest on those who would impose them. For example, in spite of repeated assertions and worries, there is no evidence at all that libraries have cancelled journal subscriptions because of publicly accessible funded research articles. Before such an assertion is allowed to impose an unnecessary burden on public access, evidence should be demanded of actual harm. Where no such evidence can be found, immediate public access should be the norm, because this is the best way to foster innovation, competition, economic growth and scientific progress.

These comments are submitted on behalf of the Duke University Libraries by:

Deborah Jakubs

Rita DiGiallonardo Holloway University Librarian & Vice Provost for Library Affairs

Kevin L. Smith

Director of Scholarly Communications

Paolo Mangiafico

Director of Digital Information Strategy

⁸ Committee on the U.S. Commitment to Global Health, Board on Global Health, "The U.S. commitment to global health : recommendations for the public and private sectors , " p.255. Available at http://www.nap.edu/catalog.php?record_id=12642 (Accessed Dec. 7, 2011).

⁹ A list of journals with the embargoes from Highwire Press illustrates that almost no journals are embargoed for more than one year, and some for substantially shorter periods. See <http://highwire.stanford.edu/lists/freeart.dtl>.