

Date: January 12, 2012

Name: Marilyn S Billings

Email: mbillings@library.umass.edu

Affiliation: University of Massachusetts Amherst, University Libraries

City, State: Amherst, MA

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 University of Massachusetts Amherst. UMass Amherst is the flagship campus of the University of Massachusetts system, sitting on nearly 1,450-acres in the scenic Pioneer Valley of Western Massachusetts, 90 miles from Boston and 175 miles from New York City. The campus provides a rich cultural environment in a rural setting close to major urban centers.

It enrolls over 21,000 undergraduate students and over 6,000 graduate students in 86 bachelor's degree programs, six associate's, 72 master's and 50 doctoral programs in eight schools and colleges. There are over 1000 full-time instructional faculty at this public land grant university that has the education of the public as one of its key missions. UMass Amherst is one of the nation's top public research universities, listed as a Carnegie Research Extensive University. The campus attracts over \$140 million in externally sponsored research each year, demonstrating its contribution to Massachusetts's position as a technological and economic leader. Funding supports the creation of new knowledge and its translation into the technical innovations and scholarly works that create opportunity for students, faculty and the public. The University Libraries is the largest public academic research library in Massachusetts. It has led conversations about open access to scholarly and creative works since 2001, actively engaged in promoting conversations through workshops, Faculty Senate debates, and providing solutions to open access through its digital repository ScholarWorks @ UMass Amherst.

The University of Massachusetts Amherst is a member of the Boston Library Consortium (BLC), the Association of Research Libraries (ARL), and is an affiliate member of the Coalition of Open Access Policy Institutions (COAPI) so many of these comments are reflections of comments provided by those entities that we want to reinforce as critical to this institution and its values. UMass Amherst fully subscribes to the principle that taxpayers are entitled to access the results of publicly-funded research, research funded by their tax dollars, immediately, and that taxpayers are entitled to fully reuse those results. The current NIH Public Access Policy, implemented in 2008, applies to the results of approximately one-third of all federally funded scientific research, and a significant amount of the research taking place at UMass Amherst. The NIH policy, while it is not without limitations, has been enormously successful in opening the results of NIH research to a broader audience - to the benefit of science and the general public. There is an urgent need for the federal government to adopt a comprehensive public access policy approach applicable to all major research funding agencies, one that would both extend and improve upon the current NIH policy. UMass Amherst agrees with the COAPI recommendation for the federal government to develop a policy framework that 1) is as uniform as possible for all agencies, 2) is mandatory for all researchers funded in whole or in part by those agencies, 3) results in rapid and open access to the results of peer-reviewed, government-funded research, and 4) allows flexible rights of reuse.

(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:

Successful development of markets related to access and analysis of government-funded peer-reviewed publications depends in large part on the speed with which research information is made available and the terms under which it can be used. The combination of rapid public access and liberal reuse rights will drive software development that facilitates new types of information discovery and tools for research.

It will create the capacity for new information-based business models that draw on the innovations in information technology, such as the semantic web, which fosters sharing and reuse of information across applications and community boundaries. Full open access in this sense will also foster commercialization of products that increase access to and awareness of specialized research information.

All of these potential capacities will be reduced to the extent that access is delayed through embargoes or that reuse rights are limited unnecessarily.

Text mining, data mining, other forms of information computation, and the creation of derivative works are examples of new research and information dissemination capacities that can be enabled through appropriate reuse rights. For example, the University of Massachusetts Center for Intelligent Information Retrieval (CIIR) is one of the leading research groups working in the areas of information retrieval and information extraction. The CIIR studies and develops tools that provide effective and efficient access to large networks of heterogeneous, multimedia information. In addition to commercial applications, such tools could also be valuable to funding agencies by allowing them to monitor research developments in specific fields as part of the process of setting funding priorities.

A broader federal public access policy framework will also foster the continued development of open access journals (which now number more than 7,000 titles) and the transition of traditional publishing to open access business models ? again to the benefit of science, economic development, and public welfare. Commercial firms ? both new firms such as Hindawi and existing ones such as Springer ? are clearly realizing the economic benefits of open access through the creation of profitable new journals that follow open access business models. Nonprofit publishers are also experimenting with open access publishing and thereby extending the reach of the research they disseminate. The growth of publicly accessible research information will encourage scholarly publishers (both nonprofit and for-profit) to transition to open access in ways that meet both their scholarly missions and their economic interests. A broader federal public access policy framework will thus both add to and encourage the continued growth of openly accessible research information.

Numerous studies have demonstrated that openly accessible research information reaches wider audiences and produces more citations than research published under access restrictions. Recent studies are also showing that openly accessible research produces more diversity in follow-on research. It encourages contributions by participants who would have had no opportunity to contribute in an environment with access controls. It thus increases the potential for innovation and the interdisciplinary application of research through a larger pool of participants.

A government-wide public access policy or policies can be implemented by leveraging existing infrastructure in ways that minimize duplication of effort. The investments in software and other

resources that already support NIH's PubMed Central and similar repositories can be utilized by other agencies either individually or in a federated model. A comprehensive federal public access policy framework will have the added benefit of increasing the effectiveness of government research funding. One of the primary motivations of the NIH policy was improved documentation of the outcomes of sponsored research. A comprehensive federal policy will bring that benefit to all of the major scientific research funding agencies. It will also provide congressional appropriators and authorizers better information to assess the value of existing expenditures and better target strategic funding priorities. It will thus increase agency accountability and support informed, transparent, and evidence-based budget and policy decision-making in accordance with the Obama administration's emphasis on open government. 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. If the goals of agency policies are to foster the development of science, encourage economic growth, and serve the public's interests in the broadest sense, then it will be important to construct the licensing framework for the policies according to principles that will facilitate those goals. Doing that requires no change in copyright law. It is only necessary to structure the licenses that authors grant to the agencies (as a condition of their funding) and the licenses that the agencies grant to the public in ways that facilitate both access to and maximum reuse of research information. A Creative Commons attribution license is an example of a license that would fulfill those purposes. Such a license would allow authors to receive full credit for their works while also creating great flexibility in terms of how their works can be used by others. 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. Licenses that allow only for access to research information but not subsequent reuse or redistribution to colleagues are unnecessarily restrictive. Unlike the NIH policy, systematic downloading of articles should be allowed in order to facilitate flexibility in terms of reuse, for example, by programs that compute on the textual corpus. Since the licensing framework for the agency policies would be non-exclusive, authors would remain in a position to transfer appropriate rights to publishers. Like the NIH policy, agency policies should be mandatory, with authors required to deposit their final (post-peer-review) manuscripts in publicly accessible repositories. In view of that, publisher transfer of rights agreements for federally funded research articles could not be structured in ways that conflict with the licenses that researchers grant to the agencies. Publisher economic interests can be protected by brief embargo periods during which the

use of the research information would be governed either by fair use under copyright for journals in print form or ? in the case of electronic journals ? by the provisions of license agreements. 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. Metadata standards would include a full citation to the publisher copy of record. Such a policy framework would balance the needs and interests of research authors, agencies, publishers, and the general public.

(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 University of Massachusetts Amherst believes that a centralized or federated approach managed by the federal government is the most appropriate and effective strategy for ensuring interoperability as well as effective search mechanisms and analytic tools. Federally managed approaches are also the most feasible way to facilitate new research capabilities related to reuse (such as text and data mining, creation of derivative works, information discovery tools, and commercialization of products that increase access to and awareness of specialized research information). Even with carefully crafted regulatory requirements, it is clearly more difficult to establish and maintain such capabilities under a decentralized framework that includes partners outside the federal government.

The federal government has a long-term interest in making the results of its funded research permanently available and 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 federal government to ensure that the long-term preservation of and access to these resources is undertaken and with appropriate use rights for the federal government and users alike. It is the only entity that has the capacity to make the full corpus of federally funded works publicly accessible, to establish and enforce standards of interoperability that ensure search access across repositories, and to establish and maintain an infrastructure that will allow new services and products to be built from publicly funded information.

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. 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.

Primary reliance on a federal government role does not preclude private or third parties from participating in a decentralized approach. However, any decentralized approach that involves entities outside the federal government, whether public or private, would need to provide all of the capacities

described above ? public access, interoperability, search functionality across repositories, adherence to standards, long-term archiving and preservation, openness and accountability, and the potential for creative reuse for research and commercial purposes. In addition, clearly delineated roles and responsibilities will be key. If the federal government found that a decentralized approach was feasible and decided to rely on it heavily, then government agencies should maintain mirrored and accessible versions of the decentralized repositories in order to protect the public's investment and ensure accountability. 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. The federal government's stewardship over this valuable public good is critical.

(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:

Academic research libraries have developed extensive experience and expertise in creating and managing digital archives designed for long-term preservation and access. Examples include arXiv (now managed by the Cornell University Libraries), the digital repositories of several research universities such as ScholarWorks @ UMass Amherst, and the HathiTrust, a major partnership of research libraries and research institutions that is designed to preserve digital books and broader cultural heritage. Given their expertise and focus on long-term preservation and access, research libraries could be important consultants in the development and implementation of federal, interagency and public/private partnerships in a public access policy. Some research universities could also partner with federal agencies to develop repositories for specific subject areas.

For example, UMass Amherst Libraries are engaged in the development of a subject repository for Nano manufacturing (InterNano) and received an NSF grant three years ago to develop a beta subject repository for materials in the ethical conduct of research in the sciences and social sciences (ESENCE). In addition, we note that some academic and research institutions have partnered with research funders to provide their permanent archives.

Publishers could be encouraged to participate in public-private partnerships by voluntarily providing the final published versions of articles after limited embargo periods that ensure their subscriptions and licensing revenues. However, given their focus on immediate income and the fact that they tend not to have long-term time horizons, commercial publishing firms in particular should not be relied upon solely for digital archiving. It should be obvious that long-term archiving and public access will be made much more difficult when corporate acquisitions, mergers, or business failures occur. For that reason, publishers should provide archiving and public access for the results of federally funded research only if the publishers' sites are mirrored by sites maintained by the federal government or by institutions that provide greater certainty of long-term preservation and access, such as research institutions. Publishers would also have to be able to comply with detailed rules for user interface, access formats, and interoperability.

(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:

The development of interoperable search, discovery, and analysis capacity across disciplines and archives depends on the creation of carefully crafted metadata standards that are implemented for all archives containing the results of federally funded research. It is critical that metadata be both machine-readable and machine-interoperable if agency policies are to realize their full potential. Metadata standards for archives should be designed to facilitate the functions of use, reuse, and analysis described above.

Federal agencies, through their public access policies, are best positioned to ensure the creation of metadata standards that will meet the functional goals of their policies. The research library community, including the Library of Congress and organizations such as OCLC, has developed a variety of metadata standards that have been endorsed by standards organizations (NISO, ISO, etc.). These can be drawn upon in developing a broad federal metadata specification.

The specification should support multiple metadata standards in order to develop metadata that is as rich as possible. Some of the primary goals of the specification (along with examples of related standards) would be to: 1) provide institutional information for published sources (grant IDs, funding organization, Institutional Identifier, etc.), 2) provide descriptive information for both the repository and published versions (Dublin Core, ORCID), 3) support searching through keywords as well as controlled vocabulary schema appropriate to disciplines, 4) incorporate abstracts, 5) facilitate full text searching and web crawling, 6) support metadata harvesting (OAI-PMH), 7) establish relationships through semantic web standards (RDF), 8) support usage tracking (COUNTER), 9) support description of related data (DataCite Metadata Schema), 10) support data exchange standards (JSON), and 11) document intellectual property rights.

It's especially important for metadata to support the capacity for machines to access and analyze both the publications themselves and the underlying data that support them in those instances where that data can be made openly accessible.

(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:

The benefits of public access policies to taxpayers will be realized to the extent that publicly funded research results are made openly accessible. The history of the development of the NIH Public Access Policy demonstrates conclusively that a broader federal public access policy (or policies) must be mandatory. The rate of compliance with the NIH policy increased dramatically following the end of the voluntary policy and the adoption of the current mandatory policy.

Average manuscript submissions have grown from approximately 1,000 per month prior to April 2008 (the date of adoption) to current levels that are well over 5,000 per month (for the most recent twelve-month period). See: <http://www.nihms.nih.gov/stats/> A broader federal policy must be consistent across all agencies in its requirements and mandates. Uniform requirements and procedures across all agencies will reduce burdens on researchers (who often hold grants from multiple agencies) and on the

institutions that support their compliance. Uniformity will reduce complexity and that in turn will reduce the time needed to educate researchers about policy requirements, to deposit articles, and to deal with deposit and compliance problems. Uniformity will also work to increase compliance rates. Publisher interests, for example those related to embargo periods and any deposit of final published versions of articles, are also best served by a uniform approach. Procedures should include standard criteria for what should be deposited as well as clear instructions for the deposit process. Existing grant management systems should also be integrated into the deposit process to facilitate agency and public accountability.

(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:

Yes, definitely. There are other important types of scholarly communications beyond the peer-reviewed scholarly journal articles.

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. At the University of Massachusetts Amherst, these policies and procedures were developed in coordination with the Graduate School, the Faculty Senate Graduate Council, and the University Libraries. Open access ETDs have been captured in the ScholarWorks digital repository since 2008. It should be noted that, 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:

Advances in science and other scholarly disciplines build upon prior knowledge and the sharing of information. The scientific, economic, and public benefits of providing access to the return on our nation's investment in research diminish to the extent that access is delayed or denied. Immediate access at the time of publication is therefore ideal in terms of overall policy goals. 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>). It is important to note that some publishers who have expressed concern in the past that public access would result in loss of subscription

revenue have changed both their views and their practices. If it is demonstrated through empirical evidence that embargoes are necessary, the University of Massachusetts agrees with members of COAPI that a uniform embargo period of six months or less should apply across all funding agencies. Such an approach has the benefits related to consistency across disciplines and would speed research access while also taking into account publisher interests. It is also 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/>). In addition, there is no evidence to support that academic and research libraries either have considered ? or would in the future consider ? public access to federally funded research to be an adequate substitute for journal subscriptions or licenses.

If a decision is made to adopt different embargo periods for individual disciplines or sub-disciplines, shorter embargo periods (less than six months, for example) should apply to rapidly changing fields and those where research results often lead directly to commercialization.

We would emphasize that the burden of proof for the need for embargoes should rest on those who believe they are necessary. The benefits of public access are clear. In the absence of empirical evidence clearly demonstrating the need for embargoes, immediate public access should be the norm, since it is the best way to foster innovation, competition, economic growth and scientific progress.

Please direct any questions about this response to:
Marilyn Billings
Scholarly Communication Librarian
University of Massachusetts Amherst