From: Ginny Steel
Date: January 9, 2012
To: publicaccess@ostp.gov
Subject: Response to RFI on open access

To Whom It May Concern,

I am writing in my capacity as the Chair of the University of California Council of University Librarians to submit the attached comments in response to the RFI on open access issued by the Office of Science and Technology Policy in late 2011. Collectively, the UC libraries make up the largest research/academic library in the world, with over 35 million volumes in our holdings as well as significant digital collections. The 10 campuses and the California Digital Library work together to expand the scope of our collections, improve access to information, and develop alternative modes of scholarly communication so that all faculty, students, and staff have access to the resources they need to support their teaching, learning, research, and service. One of the primary goals for UC's Council of University Librarians in 2012-15 is to support efforts to change the current, unsustainable models of scholarly communication that are having a calamitous and wide-reaching effect on academic library budgets and that are limiting access to important research. In light of our serious concern that there be widespread public access to peer-reviewed scholarly publications resulting from federally funded research, we offer the attached recommendations.

Sincerely,

Ginny Steel
University Librarian and
Chair, UC Council of University Librarians
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064
vsteel@ucsc.edu
831 459 2076
831 459 8206 fax

To: Office of Science and Technology Policy (publicaccess@ostp.gov)
From: University of California Libraries
Subject: Response to the OSTP RFI on Public Access to Peer Reviewed Scholarly Publications
Dec. 21, 2012

The University of California's Academic Senate has requested that the University of California system "actively encourage open access to publications by promoting national legislation and policies by federal funding agencies that support open access." [1] The Council of University Librarians is pleased to submit the following response to the OSTP Request for Information on Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research.

To preface our response, we quote from the University's systemwide faculty committee on Libraries and Scholarly Communication, which wrote in May, 2009:
"Scholars at the University of California have a vested interest in ensuring that their work reaches the widest possible audience, including members of the public whose tax dollars support much of the University's research enterprise."

Responses

(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?

Providing, open, unrestricted access to federally funded research will help to grow the economy as individuals, institutions, and companies use the freely available material to create new services and products, whether commercial or not. Faster access to scholarly output means that new constructions and ideas can be formulated at a more rapid pace, thus speeding the launch of new products into the market. Open access also enables serendipitous discoveries by readers and researchers from different fields and remote locations, leading to new ways of interpreting material and new uses for scientific (and other) discoveries. This, in turn, may further increase the reach of and market for scholarly publications.

Studies have shown that offering increased access to federally funded research brings a significant return on investment,[2] including accelerating and widening opportunities for adoption of research findings, positively impacting the quality of services offered in various fields, and the potential for new industries to emerge. The NIH public access policy provides a real-life example of how a public access policy can work.[3]

The University of California receives federal grants and research funding in fields such as life and health sciences, physical sciences, and agriculture, and offers numerous examples of the economic potential of scholarly information.[4] For example:

- UC researchers produce, on average, four new inventions per day.
- In California alone, more than 1,000 R&D-intensive companies use UC research in their work every day.
- Since 1976, 461 startup companies have been formed with UC inventions. UC's medical centers perform hundreds of clinical trials every year, resulting in new drugs and disease treatments.
- Research funding leads to a well-trained workforce and increased employment as researchers work with graduate students, train undergraduates, and hire staff.

(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?

Current copyright, intellectual property, and contract laws offer a plethora of possible arrangements for publishers, scientists, authors, and other stakeholders involved with publication and dissemination of peer-reviewed scholarly publications from federally funded research. The current business model for the vast majority of scholarly peer-reviewed publishing is for scholars to perform the research that underpins the articles, write the articles for submission to scholarly journals, and then review articles and work by colleagues for those journals. Following all of that effort for no monetary compensation, publishers require those same authors to transfer their copyrights, which then locks up the content and prevents the results of publicly funded research from being made broadly available to the public that funded it. Research libraries, not publishers, provide the aggregate, long-term stewardship and preservation of the results of publicly funded research. Insuring broad public access to and long-term preservation of federally funded research and results through a federal requirement or provision that also preserves the publisher's right to monetize the content would be a great improvement in this arena.

Another possible step would be a simple low-cost or free registration requirement and process for documenting copyright ownership that would allow for more transparent and seamless management of copyright. Current copyright law does not require formal registration before copyright is affixed, and adding a registration requirement would make it much more likely that publicly funded publications would all be made accessible and preserved. Currently, the best protection for all stakeholders may be to encourage (or mandate) immediate open access upon publication under a CC-BY license, thus ensuring that the economic foundation of the publication process is predicated on open access rather than in conflict with it, producing a far better alignment of public and private interests. Use of CC-BY

licenses would facilitate appropriate attribution and credit, which are critical to scientists and to preserving the formal record of research.

Any new policies or legislation should be sure not to increase the intellectual policy protections afforded under the US Copyright Law that would make it more difficult to insure public access and preservation of publicly funded research and publications. In order to further "promote the progress of science," scientific discoveries and the publications that result from them that are publicly funded should not be locked up so that the majority of Americans have no access. UC supports the concept of streamlining regulations that govern creative expression and research publications to insure new discoveries can be found and utilized. Additional copyright and intellectual property protection would only create unnecessary burdens that do not provide meaningful protection and delay and thwart important new research, discoveries and innovations that can be useful to the public that funded the research. The "fair use" exemption of the Copyright Law provides few options in this age of online resources that are locked up behind firewalls and pay-per-view.

(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 question of whether to centralize or decentralize the management of public access to peer-reviewed scholarly publications that result from federally funded research is a complex one with technical, cultural and financial implications. The argument in favor of <u>centralizing</u> management of this work includes the following *technical* considerations:

i. Ease in managing compliance auditing
ii. Ease in building new tools and services on top of the repository (for submission, for access)
iii. No need to develop technical solutions for edge cases where content might otherwise need to live in multiple places (i.e., interdisciplinary work, co- funded work, etc.)
iv. Control for consistent metadata. A single workflow can ensure that comparable metadata is collected for all publications submitted to the system, supporting better discoverability, refined searching, content grouping, etc.

The argument in favor of centralizing the management of these publications also includes the following *cultural and financial* considerations:

i. A centralized repository is likely to have a single submission workflow, making it easier for individuals, campuses, publishers to comply with federal deposit mandates.
ii. A centralized repository will also provide a consistent access environment for publications.

iii. The repository itself is likely to become a known research destination, with comprehensive coverage of fields of study.

iv. A sustainability plan can be developed which, if successful, will protect access to all content within the repository equally – an advantage over a distributed model where the financial health of the local repositories could be variable.

There are, however, a few technical considerations in favor of decentralization of the management of these publications.

i. A single, centralized repository will have a significant technical/resource burden in one location. Should that burden exceed available resources, the repository could be at risk – and the liabilities associated with that risk are greater when there is a single point of access for all materials.

ii. Centralized management also typically results in less capacity for customization at the local level. The one-size-fits-all model will support scalability, but will potentially limit the ability of the repository to develop distinct tools and services for distinct bodies of research.

iii. Challenges around search and discovery are more significant in vast collections. Particular attention would need to be paid to this technical challenge to avoid compromising the discoverability of any research deposited in the repository.

Regardless of the decision to centralize or decentralize management of publicly funded publications, any solution must emphasize robust discovery, access to and preservation of this research. All potential repositories must support access and use conditions that enable robust use by all interested communities – including the ability to layer services, products, etc. on top of this publicly funded research. It is also crucial that the repository infrastructure includes highly developed preservation and curation services, to ensure enduring access to the research regardless of the vicissitudes of local economic and technical environments. Third party providers might prove strong partners in these regards. However, the federal government should retain the right, regardless of repository location/technical platform, to archive and distribute publicly funded articles.

(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 HathiTrust digital archives (www.hathitrust.org) is a partnership of major research institutions and libraries who have come together to ensure that cultural history is preserved into the future by collecting, organizing, preserving and sharing the record of human knowledge. Among other goals, the HathiTrust offers preservation of digitized works from universities and other institutions. The partners include public and private institutions, and the majority of the files were digitized in partnership with Google. HathiTrust also partners with publishers. For example, Duke University Press is working with HathiTrust to make a large number of its backlist titles freely available using files digitized by Google. With permission from rights holders, the books will be made available under a Creative Commons

noncommercial license, so that anyone may read and use them for nonmonetary purposes, while Duke will be able to offer print-on-demand copies for sale. It is estimated that as many as a thousand titles could be made openly accessible under this arrangement, which is intended to be a model for other publishers who want to widen access to their material.[5]

The NLMPLus semantic search engine (http://nlmplus.com/) was developed by a start-up company called WebLib that specializes in creating search and knowledge discovery tools for the web. NLMPlus searches sixty National Library of Medicine databases to find relevant articles, including 1.6 million PubMed Review articles.

Organizations such as arXiv, JSTOR, Portico, and CLOCKSS have long-standing relationships with publishers to archive content which could conceivably be leveraged to provide access to that archived content. The JSTOR "Data For Research" service (http://dfr.jstor.org/) aggregates 6.4 million records from content archived in the JSTOR repository for content mining purposes. Such a service could be more fully developed to incorporate additional content. Ensuring that content is openly accessible would allow additional similar services to be developed and flourish, serving both general and specialized constituencies.

Services from PubMed and PubMed Central to Mendeley to Google Scholar demonstrate the power of aggregating research information in ways that enhance researchers' ability to locate, access, and work with a wide array of research content; enabling full access to content through such services will speed innovation and research.

(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?

In order to encourage interoperable search, discovery and analysis and timely deposit, this system must make the work of submitting metadata and related content extremely easy and as low-cost as possible. The attainment of some critical goals would help achieve this, including the following:

- The system must have a clearly published metadata schema that accommodates the essential bibliographic elements of scholarly published content. At this point the majority of content covered under this act is likely to fall into standard genres, i.e. monograph-like objects or journal articles. However, new genres are emerging (for instance, data sets) and those materials will need to be accounted for as well. Because these different genres of material have different core metadata elements, any schema(s) that are adopted will need to be flexible and not deeply mapped to the practices of any discipline subset. It will also be

beneficial to turn to schemas already in use, in order to take advantage of existing workflows, documentation, knowledge bases, etc.

With the above concerns in mind, an excellent starting point would be the NLM suite of schemas (http://dtd.nlm.nih.gov/) that together cover a wide array of scholarly material formats, from non-peer reviewed product reviews to formally published monographs. Many publishers and repositories from across a wide variety of disciplines use these standards already as a relatively efficient medium of data exchange. While local metadata formats may be entirely different, the NLM standards are clear and well-documented, thus are relatively reasonable as transformation targets.

- Content deposits should include both the source's original metadata record and a transformation that adheres to the published standard(s) described above. Supplying original metadata records ensures that unique metadata meaningful to content from that source will be retained, and could be surfaced to those interested in materials from that source. For instance, controlled vocabularies relevant to a particular discipline but not meaningful in a more general display system, could still be available via these records and could potentially be a source for a third-party to develop a service based on them.

- Data should be regularly exposed in a variety of ways, including a web interface for people, an OAI-PMH interface, a RESTful API, etc. New means for making content searchable should be folded in as they develop and are adopted.

- Unique identifiers for individuals (ORCIDs, once they become available) and publications (ARKs, DOIs, Handles) should be used when they are submitted with records, and should be added to records that lack them.

(6) How can Federal agencies that fund science maximize the benefit of public access policies to U.S. taxpayers, and their investment in the peer reviewed literature, while minimizing burden and costs for stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?

Federal agencies can implement consistent open access policies for the research results generated from their grant funding. Immediate public access to scientific results will maximize funders' investments by exposing the results to the widest audience possible. New policies should minimize the burden of compliance by having common standards and protocols to follow. There are existing protocols, for example, for depositing manuscripts into repositories, and the NIH/PubMed requirement also provides procedures that could be replicated by other agencies.

Direct OA publishing, supported via grant funds, is perhaps the most straightforward and least burdensome mechanism by which to implement public access for all stakeholders, since it eliminates the need for secondary deposit mechanisms of refereed articles alongside the normal publishing stream and also provides for funding of research publication at a time when such funding is at risk in many quarters. Deposit in a consistent repository or set of interoperable repositories is also desirable; however, with

direct OA publishing, articles could be harvested into repositories using automated means, without cumbersome deposit mechanisms.

(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?

Other materials that have been funded by federal funds should surely be considered for public access, but different types of materials also have other issues that would need to be considered, such as varying conventions for intellectual property (e.g., book chapters are often "works made for hire,"), digitization or lack thereof (print publications that would need be scanned to be deposited in a repository), and immediacy of need from the consumer's perspective (some material might be considered more crucial than others). A single policy most likely cannot cover all of the types of materials that might result from federally funded research.

(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?

Very limited embargoes on content that is not published as open access is generally sufficient to protect publishers. It is well established that making preprints freely available in the arXiv repository has not had a negative impact on the American Physical Society (APS) or its British counterpart, the Institute of Physics (IOP).[6] In addition, most journals on HighWire Press release their content after 12 months, as do any number of other leading journals (e.g. journals from Cell Press, some journals published by Nature Publishing Group, etc.), and subscriptions have not been affected. Perhaps the most widely known example of appropriate embargo periods (and the lack of impact on the marketplace for scholarly journals) is the NIH public access policy, which states that the results of research funded by the NIH must be available no more than twelve months after publication. A number of journals have experimented with much shorter embargo periods, without ill effect (for example, *Molecular Biology of the Cell* releases its content after two months). These examples illustrate the real viability of a federal open-access requirement that will help ensure more rapid access to vitally needed research results. Some attention may need to be paid to humanities and similar disciplines with extended periods between issues and a longer citation half-life. Shorter embargo periods are obviously preferable. The interests of the American people must guide this decision; in any analysis, the long-term public good must outweigh short-term commercial interests.

In a 2005 study on author self-archiving, Tim Berners-Lee, et. al, concluded: "All objective evidence from the past decade and a half of self-archiving, however, shows that self-archiving can and does co-exist peacefully with journals while greatly enhancing both author/article and journal impact, to the benefit

of both. Journal publishers should not be trying to delay and block self-archiving policy; they should be collaborating with the research community on ways to share its vast benefits."[7]

[1] Academic Council Chair Mary Croughan to UC President Mark Yudof, June 16, 2009, and attachment: <http://www.universityofcalifornia.edu/senate/reports/MC_Yudof_open%20access%20FINAL.pdf>

[2] John Houghton and Peter Sheehan, "The Economic Impact of Enhanced Access to Research Findings," CSES Working Paper No. 23, July 2006 <www.cfses.com/documents/wp23.pdf>

[3] NIH Public Access Policy <http://publicaccess.nih.gov/policy.htm>

[4] UC for California, specifically, "UC Impacts" <http://www.ucforcalifornia.org/uc4ca/home/resources> and Economic & Planning Systems Inc.'s September 2011 study <http://www.universityofcalifornia.edu/news/article/26216>

[5] "HathiTrust and Google Will Help Duke U. Press Digitize a Trove of Older Titles," Chronicle of Higher Education, Sept. 12, 2011. <http://chronicle.com/article/HathiTrustGoogle-Will/128949>

[6] Open Access Self-Archiving: An introduction. Technical Report, JISC, HEFCE. Swann, Alma. <http://eprints.ecs.soton.ac.uk/11006/>

[7] Berners-Lee, T., De Roure, D., Harnad, S. and Shadbolt, N. Journal publishing and author selfarchiving: Peaceful Co-Existence and Fruitful Collaboration. <http://eprints.ecs.soton.ac.uk/11160/>