

By Electronic Submission to:

publicaccess@ostp.gov

Ted Wackler, Chief of Staff
Office of Science and Technology Policy
Executive Office of the President
725 17th Street Room 5228
Washington, DC 20502
Phone: (202) 456-7116
Fax: (202) 456-6021

December 29, 2011

Dear Mr. Wackler:

The undersigned researchers of the Berkeley Digital Library Copyright Project¹ respectfully submit these comments in response to the Office of Science and Technology Policy's Requests for Information regarding public access to peer reviewed scholarly publications from federally funded scientific research ("RFI"), dated November 4, 2011.²

The Berkeley Digital Library Project is a grant-funded project that aims to investigate copyright obstacles faced by libraries and other like-minded organizations in their efforts to realize the full potential of present and future digital library initiatives. Our efforts are concentrated on both the obstacles themselves and the range of possible legal, technological, social, and market-based solutions to overcome them. Among other issues, we are specifically examining challenges with respect to orphan works, library privileges, digital lending, metadata ownership, licensing, and the ways that those issues impact the creation of a national digital public library. Principal investigators for the project are Berkeley Law professors Pamela Samuelson, Jason Schultz, and Jennifer Urban. David Hansen is the project's full-time research fellow.

The comments below explain the impact that a federal access policy would have on the creation of a national digital public library, and the broader legal issues that should be considered to ensure that such a policy aligns—or at a minimum, does not conflict with—a national digital public library initiative.

¹ UC Berkeley affiliation is provided for identification purposes only.

² 76 Fed. Reg. 68518 (Nov. 4, 2011).



I. A National Digital Public Library and Aggregated Access

Serious efforts are underway to develop a national digital public library. In October 2011 the Digital Public Library of America (“DPLA”) launched, from the National Archives in Washington, D.C., efforts to take “concrete steps toward the realization of a large-scale digital public library that will make the cultural and scientific record available to all.”³ That project has the ambitious goal of bringing together and further building upon the digital resources already made available by the likes of the Library of Congress, the HathiTrust, the Internet Archive, and the numerous universities and research organizations that currently maintain institutional repositories of digital copies of works produced from scientific research. These efforts are designed to provide access within the United States on a level comparable to that enabled by initiatives in other parts of the world, such as in the European Union (“EU”), which is already well on its way to developing an EU-wide digital public library (“Europeana”), that currently provides access to millions of items that represent Europe’s cultural and scientific record.⁴

In essence, the DPLA and related projects aim to provide widespread aggregated access to our nation’s creative and scientific output. Because federal funding plays a large role in the creation of that output, it is important to consider how a federally funded access policy might fit into an aggregated system of national digital access. In particular, because digital libraries are important in both the creation and dissemination of federally funded scientific research, they are stakeholders who have an intense interest in the intellectual property structure of such a policy. The comments below outline principles that should be considered in creating that policy. These comments are responsive to questions (2) and (3) of the RFI regarding:

- (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?; and
- (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?

II. Intellectual Property Principles For National Digital Access

A federal open access policy for peer reviewed scholarly publications would greatly further the movements described above toward widespread national digital access. The benefits of such access are myriad, and can be realized without jeopardizing the system of incentives upon which the current copyright system operates. These benefits and the ways that they complement (and not harm) the interests of scientists and researchers are well explained in the comments submitted by others (e.g., the many university libraries also replying to this RFI). Although almost any policy that increases online access to publications resulting from federally-funded research would be beneficial, a few principles can be discerned to extract the greatest value from such a policy by enabling dissemination through a national digital public library. These comments outline three principles in

³ *About*, DIGITAL PUBLIC LIBRARY OF AMERICA, <http://dp.la/about/> (last visited Dec. 8, 2011).

⁴ See *About Us*, EUROPEANA, <http://pro.europeana.eu/web/guest/about> (last visited Dec. 8, 2011).

particular—consistency, discoverability, and extensibility—that would focus the intellectual property framework of a federal policy toward a greater impact through aggregated online access via systems such as the DPLA.

a. Consistency

A federal policy should be consistent in terms of the intellectual property rights and licenses that it adopts. While such a policy need not espouse a rule of complete uniformity across all federally-funded projects (indeed, that may be unwise as incentives for creation differ among academic disciplines), it should adhere to a principle of consistency that will ensure that the rights in scholarly works produced from federal funding are subject to one of only a few licenses, and that those licenses remain consistent over time.

One of the greatest problems in aggregating digital access to current scholarly peer-reviewed publications is the wide range of rights models that exist among academic publishers. Indeed, one ongoing attempt to catalog the current rights situation by the University of Nottingham has resulted in a list of over 1000 different academic publisher contracts, each with varied terms which dictate the ways in which authors, their institutions, and others may use published works online.⁵ Those policies vary along several lines, including the version of the article that can be posted freely online, the embargo period before such a use may be made, and the types of subsequent uses to which the article may be put. At the academic institution level, this variety of rights has proved a challenge in developing institution-wide repositories for open access to research outputs.⁶ Developing a nationwide repository would only exacerbate that challenge, but a consistent federal policy could help simplify the situation considerably, and could allow projects like the DPLA to more efficiently mark items for inclusion in its digital collections.

In a similar vein, the terms upon which access to these works are made should be compatible, as far as is possible, with existing access-oriented licenses. CreativeCommons, for example, hosts a suite of oft-used licensing models—many of them applied under academic institutional access policies—which are suitable for many academic works.⁷ Adopting consistent license terms that are compatible with more broadly used language like these licenses would allow content aggregators (such as the DPLA) to more easily parse content for inclusion in its collections, to inform patrons of any use restrictions, and would allow future researchers and creators of derivative works to more easily make use of works produced under a federal open access policy.

⁵ See *Frequently Asked Questions*, SHERPA/ROMEO, <http://www.sherpa.ac.uk/romeo/faq.php?flDnum=|&mode=simple&la=en> (last visited Dec. 9, 2011).

⁶ This problem has existed for some time. For a review of these challenges from the University of Kansas—one of the first U.S. institutions to tackle this problem—see Holly Mercer & Ada Emmett, *RoMEO Green project at the University of Kansas: an experiment to encourage interest and participation among faculty and jumpstart populating the KU ScholarWorks Repository*”, in PROCEEDINGS OF THE 68TH ANNUAL MEETING OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY (ASIST), NEW ORLEANS, at 1433–41 (2005), available at <http://hdl.handle.net/1808/873>.

⁷ See *About the Licenses*, CREATIVECOMMONS, <http://creativecommons.org/licenses/> (last visited Dec. 20, 2011).

b. Discoverability

A federal access policy should also ensure that claimed intellectual property rights and related licenses that cover federally funded works are easily discoverable, both on the face of the article produced and in its machine-readable metadata. This principle complements the consistency principle noted above, as consistency in licensing is of little use if the works covered by those various licenses are not easily discernible.

An access policy which mandates the creation and attachment of consistent intellectual property metadata would go a long way toward addressing some of the most vexing intellectual property issues that have confronted digital libraries over the past decade. The problem of digital access to works that are orphaned—that is, works whose owners cannot be located, and who are therefore unable to authorize online access—has troubled the Copyright Office, Congress, and a range of digital library initiatives as they strive to provide increased access to works whose copyright status is uncertain.⁸ A proactive federal access policy that preserves copyright information would prevent a similar problem from arising with respect to federally funded scholarly research publications.

For the maintenance and discoverability of this data on a larger scale, a centralized registry of metadata—including copyright and license status—on federally funded works covered by the access policy would be beneficial. Such a registry would ease the burden of libraries and users seeking to user particular works or even large groupings of works together by providing a single search point for rights analysis. An ideal system would also host, preserve, and enable access to these works, though it would be unwise to reserve that role exclusively for a federal repository, as many other digital repositories and digital libraries are willing and experienced with doing the same.

c. Extensibility

Finally, the metadata, digital copies of articles, and associated data that are produced under a federal access policy should be structured so that their use is extensible beyond the initial repository or registry in which they deposited. Doing so would preserve the essentially public nature of these works. A federal access policy that practically restricts holding of research outputs to one agency or one private entity stifles the potential for creative new access solutions, and risks monopolizing the very works that an access policy would try to make available.

Following the principle of extensibility requires that works be made accessible and useable in formats that comport with open information data standards (e.g., linked open data).⁹ Extensibility also requires that a federal policy allow outside entities to host the materials in addition to any central repository. Although a centralized depository of these materials may be an important feature promoting access—useful for, for example, consistency over time, usage tracking and record maintenance—it is axiomatic that more, distributed copies ensure preservation better than a system which maintains only one centralized copy.¹⁰

⁸ See David Hansen, *Orphan Works: Definitional Issues* (Berkeley Digital Library Copyright Project, White Paper No. 1, 2011), <http://ssrn.com/abstract=1974614>.

⁹ See, e.g., Tim Berners-Lee, *Linked Data*, W3.ORG, <http://www.w3.org/DesignIssues/LinkedData.html> (explaining the basic structure and benefits of linked open data for connecting information on the worldwide web).

¹⁰ See, e.g., Stanford University Libraries, LOCKSS Program, <http://www.lockss.org/lockss/Home> (last visited December 9, 2011) (“LOCKSS (Lots of Copies Keep Stuff Safe), based at Stanford University

Most importantly, extensibility requires that licenses to the covered works be tied not to the particular entity hosting the works, but rather to the ultimate use of the work, which is available to all hosting entities (and users). License terms should also be permissive in the uses they allow, being careful to expand and not contract the uses to which libraries and end users may put the covered works. Existing exceptions to copyright, such as fair use or the library exceptions codified in Section 108 of the Copyright Act should not be limited by license terms. Contractual limitations on these exceptions have frustrated library efforts to increase online access for their patrons in the past,¹¹ and such limitations could similarly bind federally funded research in a way that prevents useful access and that curtails future productive uses of these works. These are precisely the type of activities that existing copyright exceptions (and in particular, fair use) are designed to foster.

Adopting a principle of extensibility would enable national digital access through an entity like the DPLA, but would also serve an important role beyond that initiative. While the DPLA currently rests on today's horizon, untold projects to improve access and usability of scientific research remain foreseeable for years to come. A federal access policy designed on a principle of extensibility would enable those, as well as current, access solution projects.

An access policy for peer-reviewed scholarly publications resulting from federally funded research that is built upon the principles outlined above would increase the reach and impact of federal research dollars. In particular, dissemination through channels such as the DPLA could maximize the impact of a federal policy by integrating that content into a platform designed to reach a broader segment of the American public.

Respectfully,

A handwritten signature in blue ink, appearing to read "David Hansen".

David Hansen
Digital Library Fellow
On behalf of David Hansen, Pamela Samuelson, Jason Schultz, and Jennifer Urban.

Libraries, is an international community initiative that provides libraries with digital preservation tools and support so that they can easily and inexpensively collect and preserve their own copies of authorized e-content.”).

¹¹ See, e.g., Stanley Wilder, *The Erosion of Fair Use Protections for Digital Scholarship*, CHARLESTON ADVISOR 57, 57–58 (2011), http://www.carli.illinois.edu/mem-serv/mem-train/1104ereslicensing/Erosion_of_fair_use.pdf.