January 6, 2012

Catherine Casserly; cathy@creativecommons.org
Timothy Vollmer; tvol@creativecommons.org
Creative Commons
Mountain View, CA

Re: OSTP Request for Information: Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research [http://federalregister.gov/a/2011-28623]

Creative Commons (CC) is pleased to submit comments to the Office of Science and Technology Policy's Request for Information (RFI) on the topic of Public Access to Peer-reviewed Scholarly Publications Resulting From Federally Funded Research. Creative Commons (http://creativecommons.org) is a 501(c)(3) U.S.-based nonprofit corporation dedicated to making it easier for people to share and build upon the work of others, consistent with the rules of copyright. CC develops legal and technical tools used by individuals, cultural, educational, and research institutions, governments, and companies worldwide to overcome barriers to sharing and innovation. Creative Commons operates globally. The international CC affiliate network consists of 100+ affiliates working in over 70 jurisdictions, and there are over 500 million CC-licensed works available on the web.

Thousands of academic researchers release journal articles, datasets and educational materials under Creative Commons licenses, allowing those materials to be easily found, accessed, and reused around the world. CC licenses offer a flexible set of permissions so that authors and publishers can release their scholarly publications on the terms they wish while ensuring that they receive attribution for their work.

We answer the specific questions laid out in the RFI below.

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
This RFI comes in response to The America COMPETES Reauthorization Act of 2010. The purpose of that law is "to invest in innovation through research and development, and to improve the competitiveness of the United States."[1] All the suggested priorities listed in Question 1 of the

[1] See http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=111_cong_public_laws&docid=f:publ358.111

RFI (e.g., "grow existing and new markets," "grow the economy," "improve the productivity of the scientific enterprise") should be explored and acted upon with the public interest in mind. The public funds tens of billions of dollars in research each year. The federal government should set up policies so that the public has immediate, cost-free, unimpeded access to the research and publications that are developed with public funds (taxes).

Scholarly articles created as a result of federally funded research should be released under full open access. Full open access policies will provide to the public immediate, free-of-cost online availability to federally funded research without restriction except that attribution be given to the source.[2]

The complete collection of articles resulting from publicly funded research needs to be made available under a full open access regime so that the maximum value of federally funded research publications can be realized. In this context, we speak of value in a broad sense, including economic, scientific, and social value. Full open access policies will maximize scientific productivity, accelerate commercial innovation, and support the public's right to access and use the materials it pays for.

The federal government can support economic efficiency of the taxpayer dollars they expend by instituting an open licensing policy. The incumbent system oftentimes results in the government essentially paying into the research chain from various angles: first, when the government pays for the development of research by awarding federal grants; second, when the government subsidizes the salaries of the academics who serve as peer-reviewers; third, when the government pays for libraries in research universities that subscribe to the journals.[3] The public should receive a better return on the investments being made for the research, development, and publication of federally funded scholarship.

Existing policies such as the NIH Public Access Policy are a good first step to increasing access to federally funded research outputs. And, as suggested during this comment period last year, the federal government should require that all grantees that receive federal research funding deposit either the final, published version of their peer-reviewed journal articles or the final electronic manuscript of such an article in a publicly available digital repository. However, the NIH Public Access Policy does not go far enough in communicating the rights for reuse that should be available to the public that paid for the development and publication of that research. Releasing the outputs of federally funded research free of cost online should be a baseline, but if downstream users (including researchers within the same domain, scholars from other disciplines, creative entrepreneurs, government employees, citizens) are unclear about their legal right to copy, amend and redistribute the federally funded research, those publications will be reused less. This will significantly diminish the potential impact of the research—and, by extension, the public's investment. With a renewed attention to clarifying available rights in

[2] Carroll, Michael. *Why Full Open Access Matters.* PLoS Biology, November 2011. Volume 9, Issue 11. Available at http://www.plosbiology.org/article/info:doi%2F10.1371%2Fjournal.pbio.1001210.

[3] Mays-Smith, Simon et al. *Three Pillars of Wisdom: the STM journal industry*. Credit Suisse First Boston. April 2004. Not available, but see http://www.earlham.edu/~peters/fos/newsletter/05-03-04.htm#creditsuisse for a synopsis.

advance, the federal government can increase the speed of scientific discoveries, promote innovation, and support new business opportunities looking to capitalize on federally funded research outputs. And as scientific researchers, creative startups, and traditional commercial players are granted the reuse rights they deserve, these and other groups can best leverage federally funded research to advance the scientific enterprise and increase U.S. competitiveness by developing new products and services.

One study showed that restricted access to gene-level research led to 30% less downstream scientific research and product development when compared to open access.[4] By extension, in at least some disciplines, scientific innovation and economic activity that is desired under the America COMPETES Act is more likely to occur on federally funded research that is made available under open access than on research that is locked down with technological protection measures, that charge for access, or that restrict use to only noncommercial purposes.

If federal agencies want to maximize the impact of publicly funded research, they should provide explicit, easy-to-understand information about the rights available to the public. The simple process of posting federally funded research articles on the web so that they can be viewed (and even downloaded) is not sufficient to promote scientific progress. Even where it is the intention that research created with public dollars be widely shared, as long as those materials are not clearly marked with information describing the rights and permissions under copyright, the use of the resources will be diminished and impact of public investment lessened. The standard means for granting permission to the public is through a Creative Commons Attribution (CC BY) license. This license is aligned with the principle of full open access because it allows rights (including commercial rights) to be communicated with the only requirement that users give credit to the rightsholder.[5]

Full open access allows for a greater number and diversity of researchers to participate in the process, and can help promote access and reuse by researchers in related fields who might not have had access to the full corpus of research because it was too expensive or because they were not granted full reuse rights. By allowing more people to interact with federally funded scholarly research, federal agencies open the door to increasing the impact of scientific discovery.[6] It has been demonstrated that reaching out to communities of researchers beyond the primary research domain helped solve one-third more problems that experienced research and development firms

[4] Williams, Heidi. *Intellectual property rights and innovation: Evidence from the human genome*, December 2009, p. 25. Available at http://deugarte.com/gomi/Williams_jmp.pdf.

[5] CC BY is a copyright license that grants permission to the public to reproduce, distribute, perform, display or adapt the licensed materials for any purpose so long as the user gives attribution to the author or as otherwise directed by the copyright holder. For more information, see https://creativecommons.org/licenses/by/3.0/

[6] There's also been growing participation in "citizen science," whereby the public is able to discover and assist with complex scientific research projects. Over 500,000 people participate in the Zooniverse (https://www.zooniverse.org/) citizen science initiative.

had been able to solve by themselves.[7] One study reasoned, "the significance of this effect may be due to the ability of 'outsiders' from relatively distant fields to see problems with fresh eyes and apply solutions that are novel to the problem domain but well known and understood by them."[8] Empowering the unexpected reader by making federally funded research outputs full open access is one method for opening the discovery horizon.[9] Another study suggests that supporting open access to research may contribute to the exploration and creation of novel research lines."[10] Making federally funded research available under full open access promotes the advancement of research from previously unforeseen angles, and communicates broad reuse rights so that economic activity and new markets can blossom.

Policies aligned with full open access should be the default for federally funded research. By doing so, the federal government maximizes their investments by encouraging competition because research is available to all to build upon, including commercial enterprises. As a result, such research outputs—and the business opportunities they generate—will be more attractive to private, venture capital, and foundation funding. Already, some philanthropic foundations have adopted open licensing policies for particular grant programs. These policies require that any grant recipient must agree to release outputs arising from foundation funding under open licenses, such as the Creative Common Attribution license.[11] If federal agencies adopt a licensing policy aligned with full open access, the outputs of federal funding can be shared, combined and repurposed with similarly licensed research, educational content and data funded by philanthropic foundations. The network effects and utility of the research increases when data and other content are legally and technically interoperable under a common licensing framework.

Full open access to federally funded research articles can be a central driver of scientific innovation and productivity. Federally funded scholarly research should be "read/write" and not simply "read only." Scientific and scholarly communities conduct research, for example, not only directly through the act of reading, but also through computational methods such as the mining of scientific literature and data. Not only would a full open access policy expand the number and diversity of human readers able to access and reuse federally funded scholarly publications, it would provide fuel for powerful new machine learning capabilities, which are

---

[7] Lakhani, et al., *The Value of Openness in Scientific Problem Solving*, October 2006, p. 2. Available at http://www.hbs.edu/research/pdf/07-050.pdf.

[8] Ibid, at 12.

[9] For example, see Kevin Smith's description of "unexpected readers" at http://blogs.library.duke.edu/scholcomm/2011/11/15/the-unexpected-reader/

[10] Murray, et al., *Of Mice and Academics: Examining the Effect of Openness on Innovation,* October 2008, p. 3. Available at http://www.hbs.edu/units/tom/seminars/2007/docs/Of%20Mice%20and%20Academics%20Stern.pdf.

[11] For example, the Shuttleworth Foundation requires that grantees apply CC BY (or allow CC BY-SA, in some cases) to the content they produce with grant funds. See http://www.shuttleworthfoundation.org/about-us/our-philosophy/open-licensing/. The Gates- and Hewlett-funded Next Generation Learning Challenges grant program stipulates that grantee outputs be licensed under CC BY. See http://nextgenlearning.org/sites/default/files/IP_Policy_105.pdf.

able to conduct novel types of computational research on huge sets of open texts. Under the current regime of restricted (or even unclear) access and reuse capabilities, such innovation and next-generation semantic text analysis is either impossible, prohibitively expensive, or fraught with potential legal liability.

Full open access makes financial sense and maximizes the impact of federal spending on research. Adopting a broad open access policy for all US federally funded scientific articles is projected to produce an estimated 5X return on investment.[12] The benefits of a public access policy similar to the existing NIH Public Access Policy would be 8X greater than costs, and the net gain of extending NIH-like policies to all U.S. federal agencies with annual extramural research budgets greater than $100 million is estimated to be about $1.5 billion.[13] The NIH public access policy has proven to be cost effective. Currently, it costs approximately $4.5 million per year to deliver the current level of access (assuming ~80,000 articles/year).[14] Thus, the annual implementation of the NIH public access policy costs 1/100th of 1% of the annual NIH budget (assuming ~$30 billion annually).

It is possible for a government-wide full open access policy to be implemented in a cost-effective manner because the federal government has already built the PubMed Central repository. In order for federally funded research outputs to align with full open access, such a repository would need to also communicate rights statements such that any scholarly publication be released with at most the requirement that attribution be made to the author (as mentioned, the standard means for communicating these rights is through a Creative Commons Attribution license). Such an addition would go a long way to making sure that the public knows how they may reuse federal funded scholarly research publications. Several universities have already integrated the technical means to support open licensing within their institutional repositories.[15] Leveraging existing infrastructure and current community best practices is increasingly important because agencies continue to face significant budget reductions.

In addition to speeding the potential for scientific discovery and research, adopting a full open access policy for the outputs of federally funded research will benefit integral government processes too. Government agencies will be able to operate more efficiently because they too will have access (and will know the permissions available to them) to the outputs of federally

---

[12] Houghton, et al., *Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs*, July 2010, p. 7-8. Available at http://www.arl.org/sparc/bm~doc/vufrpaa.pdf.

[13] Ibid.

[14] National Institutes of Health, *Analysis of Comments and Implementation of the NIH Public Access Policy*, September 2008, p. 22. Available at http://publicaccess.nih.gov/analysis_of_comments_nih_public_access_policy.pdf.

[15] Examples include MIT's DSpace (http://libraries.mit.edu/dspace-mit/index.html) and the University of Michigan's Deep Blue (http://deepblue.lib.umich.edu/).

funded research publications.[16] And, with full open access, Congress and appropriations bodies will be able to more easily see the return on investment by assessing the value of existing expenditures, and target future federal funding toward the most promising and high impact research areas.

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?

Comment 2
The current climate around intellectual property (IP) is increasingly one of ratcheting up enforcement and maximization of copyright and other IP rights. However, it's important to understand the underlying priorities of scholarly authors to determine whether the default copyright regime reflects their needs and supports their academic endeavors. The primary motivation for scientists to publish their works is "to inform others about their work."[17] Scholarly authors wish to have their research read, consumed, and cited. By adopting a full open access policy for the outputs of federally funded research, the federal government would further the fundamental goal of the scientific community.

A full open access policy for the outputs of federally funded scientific research is not incompatible with protecting the copyrights of authors and publishers. Creative Commons—and CC BY, the license aligned with full open access—is not a replacement for copyright. CC licenses operate within the current system by allowing authors and publishers to retain their copyright but mark their work with the rights they wish to communicate in advance. At the same time, CC licenses ensure that the rights holder receives credit in the manner they prefer. And, policies aligned with full open access encourage scholarly authors to share their work widely, thus increasing the ability for their work to be read and cited. One study shows that research made available via an open access repository (as opposed to a restricted access repository) experience a citation boost between 50-125%.[18]

There are now alternative publishing models for making research articles available to readers. Under a model leveraged by the Public Library of Science (PLoS), the author pays for publication and the resulting article is made available under full open access. In this way, PLoS

[16] In 2009 the Department of Justice paid about $4 million each year for access to public court records via PACER. See http://www.wired.com/threatlevel/2009/12/doj-pacer/.

[17] Hansen, et al., *Intellectual Property Experiences in the United States Scientific Community*, 2007, p. 8. Available at http://sippi.aaas.org/Pubs/SIPPI_US_IP_Survey.pdf.

[18] Furman, Jeffry L and Scott Stern, *Climbing Atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research*, National Bureau of Economic Research working paper. September 2006, p. 4. Full version not openly available. Abstract at http://www.nber.org/papers/w12523.

is exercising copyright, but simultaneously promoting openness via the use of open licensing (CC BY) instead of the default all rights reserved approach.[19]

The U.S. federal government should take into account the broad range of stakeholders in its policy-setting process, and should consider the needs and requirements of the global science and scholarly research communities. More than ever before, scientific research is no longer conducted in national silos. Digital networking, advanced communication technologies, and widespread information-sharing tools support a truly global research ecosystem where disparate researchers around the world collaborate with each other on scientific activities. The U.S. will progress most rapidly if it can take full advantage of the world's research results by urging other governments to do the same. The full value of our nation's collective investment in scientific research can only be realized if we allow these articles to be freely accessed, used, and built upon. The U.S. can participate in and even lead such efforts by adopting a full open access policy for its federally funded research outputs.

We assume the federal government is interested in widespread distribution of the non-defense research it funds because it increases the reach and impact of that research. If the federal government wishes to maximize the impact of the public's tax dollars, it should set the license and copyright terms for content created using its funds. A grant recipient may choose, or not, to accept those terms when funding is offered. The federal government might explore other options such as full open access policies, which exercise copyright law less than "all rights reserved" by leveraging open licenses to communicate rights and permissions to downstream users.

Open licensing requirements are already in place in some federal grant programs. The Department of Labor's Trade Adjustment Assistance Community College and Career Training grant program mandates that grant recipients make available any new content created with grant funds under the Creative Commons Attribution license.[20] The CC framework ensures broad access and re-use for persons wishing to utilize the federally funded research while

[19] As an example, see the PLoS Biology Terms of Use, available at http://www.plosbiology.org/static/terms.action

[20] The open licensing requirement in the Department of Labor Solicitation for Grant Applications says: "In order to further the goal of career training and education and encourage innovation in the development of new learning materials, as a condition of the receipt of a Trade Adjustment Assistance Community College and Career Training Grant ("Grant"), the Grantee will be required to license to the public (not including the Federal Government) all work created with the support of the grant ("Work") under a Creative Commons Attribution 3.0 License ("License"). This License allows subsequent users to copy, distribute, transmit and adapt the copyrighted work and requires such users to attribute the work in the manner specified by the Grantee. Notice of the License shall be affixed to the Work." Available at http://www.doleta.gov/grants/pdf/SGA-DFA-PY-10-03.pdf.

simultaneously ensuring that proper credit and important legal protections are retained.[21] CC licenses have also been upheld in copyright litigation are enforceable under copyright law.[22]

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?

Comment 3
The federal government is the appropriate entity to provide permanent stewardship and long-term access to federally funded research outputs. To ensure that the federal government is able to provide that stewardship and make federally funded scholarly research permanently available online, it must have the adequate authority to allow it to save and archive scholarly articles. This is familiar territory, as the NIH Public Access Policy already requires that grantees submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central upon acceptance for publication. Grantees (or their publishers) have had to manage copyrights properly in order to make sure that the final manuscript is submitted with the necessary authority to grant NIH permission to make it publicly available within 12 months of publication.[23] The full open access framework would require authors and publishers to ensure that the federal government has the rights to share the scholarly publications resulting from federal funds under a license that offers full open access, such as CC BY.

As mentioned previously, government stewardship of federally funded research is cost effective because the government has already developed the underlying infrastructure and related policies. However, a distributed approach could be considered as long as it can deliver a level of service that ensures that federally funded research articles are made available under full open access. Several U.S. research universities have already demonstrated they have the necessary technical, policy, and institutional stability to host distributed repositories.

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?

---

[21] There are additional features to CC licenses that grant broad access and re-use for persons wishing to utilize federally funded research outputs while ensuring that proper credit is given to federal agencies and authors sharing that research. See http://epsiplatform.eu/sites/default/files/Topic%20Report%20No%2023%20Ceative%20Commons%20and%20PSI.pdf, p. 7-10.

[22] For more information, see http://wiki.creativecommons.org/Case_Law.

[23] Mike Carroll previously explains these options in the context of managing copyrights with regard to the National Institutes of Health Public Access Policy. *Complying with the National Institutes of Health Public Access Policy: Copyright Considerations and Options*, SPARC/Science Commons/ARL. February 2008, p. 3. Available at http://www.arl.org/sparc/bm~doc/NIH_Copyright_v1.pdf.

Comment 4
Public-private partnerships should be encouraged so long as they align with the requirements of a full open access policy. This ensures that wherever researchers access/locate/read the publicly funded scholarly articles, it will be clear how they may reuse that content.

Some for-profit publishers may argue that they should host federally funded research articles published in their journals. But the long-term viability of privately managed repositories is that their long-term viability is not guaranteed. Due to the volatility in the publishing market, for-profit publishers should not be the sole gatekeepers for public access to publicly funded research. A publishing house could go out of business and access to the research articles they hold would become inaccessible to the public that paid for that very content.

Despite those concerns, if privately managed repositories can demonstrate they can meet every necessary requirement, it could encourage the development of new markets and promote innovative ways to distribute and capitalize on federally funded scholarly research.

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
Federal agencies should mandate that as a condition for receiving federal funds grantees must make available the metadata specified by the agencies. Since it is desirable for the outputs of federally funded scholarly research to be made available under the full open access regime, federal agencies should mandate that licensing information aligned with full open access be attached to each publication. Metadata should be viewed as a mechanism to enable specific actions, to proactively provide the information that researchers, machines, and unexpected readers will use to study and innovate on top of federally funded research.

By mandating that scholarly publications that arise from federally funded research be released under full open access, federal agencies ensure that downstream users will know how they may use that research. With robust metadata and thoughtful architecture for repositories (be they centralized or distributed), researchers can be assured that definitive versions of their original federally funded research papers are accessible, maintained, and archived. It will also alleviate concerns that some researchers have in making their scholarly or scientific research available for derivative use. With proper licensing and marking, the federal government promotes the correct attribution for authors and preserves the integrity of the federally funded research.[24]

[24] For more information, see a blog post on the ePSIplatform blog, "CC Tools and PSI: Supporting attribution, protecting reputation, and preserving integrity," February 2011. Available at http://epsiplatform.eu/content/cc-tools-and-psi-supporting-attribution-protecting-reputation-and-preserving-integrity.

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
Federal agencies that fund science can maximize the benefit of public access policies to U.S. taxpayers by adopting a full open access policy. For the reasons mentioned above, a full open access policy ensures the greatest level of access to the public while simultaneously making sure that the author receives credit for the work s/he creates using federal funds. For any public access policy to be effective, grantees must be clear on the requirements they must follow if they accept taxpayer money. Requirements should be consistent and clearly communicated. Full open access policies can be communicated clearly by adopting the CC BY license. The permissions and requirements of CC BY are simply and easily understood by licensors and licensees, and are also available for consumption by search engines via machine-readable metadata.

To minimize burden and costs for stakeholders, uniform language with regard to full open access policies should be used across all federal agencies. As mentioned above, there is already draft language set in place that mandates CC BY for the Department of Labor TAACCCT grant program. That language can be used as a template to communicate the legal regime under which federally funded scholarly research should be made available. Uniform deposit requirements (including licensing requirements) will make it easier and more streamlined for researchers to properly comply with federal guidelines. It will reduce complexity and cost and increase the rate of compliance overall.

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
All materials created as a result of federal funding should be made available to the public either by placing them in the public domain or making available under a CC BY license. The public should have open, unimpeded access to the content paid for with their tax dollars.

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
The public should be granted immediate access to the content of peer-reviewed scholarly publications resulting from federally funded research. Immediate access is the ideal method to optimize the scientific and commercial utility of the information contained in the articles. Especially for the scientific community, publicly funded scholarly research should be immediately available to speed the discovery of cures for diseases and promote the advancement of science.[25] Immediate access speeds up the research and development cycle, thus leading to faster development of value-added research, products and services. Immediate access can promote the creation of new and unexpected jobs in the agricultural, green energy, and biotechnology sectors, and boost interdisciplinary research collaborations and partnerships. Of course, if articles are published in an open access journal, there is no embargo period. For example, PLoS provides immediate, free access and unrestricted reuse (with attribution required under CC BY) to their peer-reviewed journals.

We thank OSTP for the opportunity to provide comments to this RFI, and we're happy to answer any other questions you may have.

Sincerely,

Catherine Casserly, CEO, Creative Commons
Timothy Vollmer, Policy Coordinator, Creative Commons

[25] The Tuberculosis Commons (TB Commons) Initiative Open Innovation Team "believes that open models can accelerate knowledge turns resulting in a faster drug development process." Content provided by end-users to the site is released into the public domain using the under the CC0 Public Domain Dedication tool. See http://www.tbcommons.org/initiative/.