

January 12, 2012

To: Office of Science and Technology Policy Executive Office of the President  
725 17th Street Room 5228  
Washington, DC 2050

From: Nathaniel Hoffman

Re: Response to the White House RFI on OA publications

I am a videogame developer who makes frequent use of government-funded research in my work. I am also active in my professional organization, the Association for Computing Machinery (ACM), including volunteering for publication-related activities such as journal paper review and conference organization.

Overall, I largely agree with Harvard's position on this issue (<http://osc.hul.harvard.edu/stp-rfi-response-january-2012>). My strongly-held opinion is that every federal agency funding non-classified research should require immediate free online access to the full-text, peer-reviewed results of that research, without any time delay. The government should also provide the means (e.g. web-accessible database archive) to provide this access, as it does with PubMedCentral. Furthermore, these means should be made available to the copyright holders of any properly peer-reviewed non-government-funded research publication on an elective basis. This is the minimum that US taxpayers deserve - full and immediate access to the research they have paid for.

The US government should also encourage other government funding bodies (e.g. the European Union) to incorporate similar mandates, and should strive for mutual agreements to automatically make the publications from each government's archive available in the others, or possibly even set up a shared archive. To reduce inconvenience to researchers, ideally such agreements would also state that submitting the work to one of these archives counts as fulfilling the mandates for all the governments in the agreement.

Given that the vast majority of peer-reviewed research is funded by some government or other, my expectation is that even the few papers not covered by this mandate would eventually be submitted to these archives, making them complete repositories of all mankind's research. This would result from pressure by the authors of this research, who would wish to maximize the availability (and thus impact) of their papers. The end result would be universal open access.

It is hard to overstate the benefits of universal open access. The reduction in costs to educational institutions (who now spend huge amounts on journal subscriptions) would be, although large, one of the least significant benefits. Far more important would be the reduction in research friction - any researcher, practitioner, student or hobbyist could immediately access any research results. Even researchers in large institutions and industry practitioners in relatively deep-pocketed companies do not currently have access to all research, since there are many publishers and professional societies, each with their own paywalls and separate non-open archives. For example, I work at a very large game company and have access to three different paid archives, and I still regularly encounter papers I do not have access to. Since the additional cost of each such paper is not negligible, I then need to weigh carefully whether the value of the paper exceeds its cost - something that is often hard to determine without reading the paper in question! This introduces a huge amount of friction and limits my

productivity when doing research and development work. How much worse must it be for small companies, self-employed people, people working in third-world countries, etc.?

This problem is especially bad for anyone doing cross-disciplinary work (for example, my own specialty, computer graphics, involves elements of computer science, optics, electronics, and others) since each discipline typically has its own set of archives. Some publishers (like the ACM) allow authors to post "preprints" (trivially different than the official published version of the paper) on their own websites, and this is indeed preferable to not allowing such. However, this is not enough; although most authors do so (to maximize availability and impact of their papers), there are always a few who do not.

Since most technology-driven businesses are based in the USA, the US would benefit from universal open access more than other countries.

However, even if this were not the case it would still be highly beneficial to the USA to institute such policies. Research is not a "zero sum game" where every benefit to one party implies a corresponding loss to another. Improvements in technology will increase productivity and economic output worldwide, benefiting the USA as it benefits other countries.

There will be scientific publishers which will claim (most likely in response to this very RFI) that these open access mandates are unfair, that they will cause economic hardship and result in job loss. This claim should be ignored, for two main reasons. One is that through a series of historical accidents, scientific publishers found themselves in a position where they extract all the value from peer-reviewed scholarly publications while contributing a negligible amount to their creation. I know full well (from personal experience as well as that of many of my colleagues) that besides the research itself (which is typically funded by government bodies and for which in any case the publishers can take no credit whatsoever), all significant parts of the process of creating a peer-reviewed paper are performed by unpaid volunteers; researchers and members of the scholarly community performing a type of community service for largely altruistic reasons.

Paper reviews, organizing paper reviews, final decisions on acceptance, are all done by unpaid volunteers. There was a time when the scientific publishers would add some value in typesetting, printing and distribution. However now that electronic distribution via the Internet is the rule and most of the typesetting is done on computer by the paper authors (as can be seen by anyone comparing an "author preprint" to the final paper - the differences are negligible).

Open access mandates may or may not adversely affect the financial situation of the scientific publishers. However, this is immaterial.

The benefits of universal open access to all sectors of the US economy, to the advancement of science and technology, indeed to the betterment of all mankind far outweighs the profits of a small group of companies which once served a valuable purpose but do so no longer.

If they manage to find some significant value that they can add to the scientific process, they will survive and even flourish despite no longer being able to continue their current rent-seeking behavior. If they do not find some way to provide significant value, then their demise need be of no concern.

Some professional organizations (like my own, the ACM) also serve as scientific publishers, in addition to their other activities. These non-profit organizations were created, and continue to exist, only to advance the science and practice of a given field of human endeavor (computing, in the case of the ACM). Unfortunately, many of them (sadly, including the ACM) have management that has grown accustomed to the revenue streams attendant upon their publishing operations, to the point that they oppose open access despite its obvious benefits.

As a member, I strongly feel that this position - which values publishing revenue over the advancement of the computing field, which is the very purpose of the ACM's existence - is proof that the ACM's management has been sadly corrupted by reliance on these revenue streams. If these revenue streams disappear as a consequence of open access mandates, the ACM will be a better organization for it, more responsive to the desires of its members and the advancement of the computing field. I expect officers of the ACM to respond to this RFI with claims that open access mandates cause the ACM damage, and that if they were extended the damage could very well render the ACM unable to continue its various beneficial activities. I have two answers to those claims - that they are almost certainly untrue, and that even if they were true it would not matter. The claims are almost certainly untrue because some combination of other revenue sources would most likely be found to make up any shortfall in publishing income. The claims do not matter because all the good that the ACM (or any other professional society, or all of them put together) do pales in comparison to the benefits of universal open access to scholarly research.

Here are my answers to the specific questions asked in this RFI:

(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?

Answer: A full and immediate open access funding mandate by the US government, combined with encouragement of other governments to follow suit, will lead to universal open access and greatly increase the productivity of the scientific enterprise. The costs are negligible, and the benefits are immense. A model similar to PubMedCentral, but covering all federally-funded research and without the 12-month delay, is the best policy.

(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?

Answer: The intellectual property interests of publishers should not be a concern - their unique contributions to the scientific process (typesetting, printing and physical distribution) are now all irrelevant. The scientific publishing enterprise can continue very well without them. The intellectual property interests of scientists are served by maximizing the distribution (and thus the impact) of their research. The intellectual property interests of Federal agencies and other stakeholders are also served by maximizing the distribution and minimizing the friction of access to research. Policies which prioritize publisher profits over scientific advancement should be avoided.

(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?

Answer: The PubMed Central model is successful and proven. The US government should maintain its own archive with a copy of all research, easily accessible and searchable. This minimizes burdens on researchers, and maximizes long-term archiving stability as well as ease of use. However, there is no reason to prohibit additional, decentralized repositories. Many authors will probably continue storing a copy of their work on their own institutional web pages - the difference being that they will no longer be forced to do so in order to keep their work accessible, as they are today. Ideally, there would eventually be multiple government archives (one for the US, one for the European Union, etc.), each of which has all the papers present in the others as well (automatically, via exchange agreements).

(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?

Explicit public-private partnerships are not needed - the government can handle archiving. Existing publisher archives can definitely continue to exist alongside if the publishers wish to keep them, but they would no longer be the sole source of research results.

(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?

I recommend to continue building upon the successful PubMed Central model in these matters. A centralized government archive, with similar metadata to that used currently in PubMed Central.

(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?

Full and immediate open-access mandates maximally benefit all stakeholders except for the (now irrelevant) publishers.

(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?

Yes, definitely.

(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?

Zero - there should be no embargo period. Any embargo period is to the detriment of all stakeholders except the (now irrelevant) publishers.