Subject: **Re: [sparc-advocacy] Call to action: 2011 White House RFI on public access (deadline Jan. 2)**

Date: November 30, 2011 3:07:48 AM EST

Dear Task Force on Public Access to Scholarly Publications: please see my responses to your questions in the Federal Register Volume 76, Number 214 (Friday, November 4, 2011) below:

(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?

To increase access, simply introduce APIs (application programming interfaces) to allow people to write programs which can easily download the data (note - I have not reviewed existing APIs). And when I say data, I mean more than descriptive data - encourage actual scientific data to be structured into articles; for example, in the medical context this means numbers in trials, results of various tests, codes for randomization procedures, etc. As far as maximizing economic growth - making this information public may not be reflected in increased GDP, since GDP is a function of spending. However, it would certainly increase public welfare and provide great benefit to human beings in general by steadily increasing the scientific literacy of the general populace. This has additional effects of spurring innovation through inventions and also potentially greatly increasing the health of the citizens, as citizens are able to bring medical research to the attention of their providers or apply it to themselves.

(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?

This question is somewhat oddly worded - both questions are asking the same thing; thus, there is no need for two questions. Also, it uses the word "conversely" incorrectly.

The intellectual property question is a tricky one. As far as I'm concerned, the efforts need to be focused on minimizing the way that large corporate publishers, which act largely as just warehousers and copyeditors but, through holding title to a journal name, extract rent from the individuals which are creating the true value: the scientists, peer reviewers, etc. The government should not permit scientists to relinquish their and the government's intellectual property rights to a journal company like Elsevier, LexisNexis, Wiley, etc.

(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 pros of decentralized approaches outweigh the cons, but the best might be a mix of the two. Obviously, a highly reliable centralized source is necessary. as the goto source. But public data, far from being unusual, is *typical* for government-funded work. And the fact that government data is public record has been enormously helpful for the public. In most important areas, data is actually public: medicine, weather, countless sciences, and even corporate finance (look at SEC filings). Even in insurance, the industry I work in, data is aggregated and much is public as it is filed with the government.

Wikipedia and open-source software provide examples of what happens when you have ultimate decentralization, although these are fairly different in that they don't get tax revenues. There are plenty of examples of open-source software allowing groups to free themselves from inflexible bureaucracies and innovate. In theory Wikipedia faces the same thing and may some day "fork". The framework could theoretically be actually structured in such a way that this government institution could be subject to the same competitive pressure.

(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?

Finance has lots of examples, and the finalization of XBRL will bring even more. Obviously, there are winnders and losers - and the existing companies are likely losers, just like the cable companies can be losers to say Netflix. On net I would expect overall administrative costs to deceline, which means lobbying companies lose jobs and revenue, while a higher percentage goes into actual research and data analysis. Not sure if this qualifies as research per se. Otherwise, I don't know - altho public government data is used by private companies all the time, often with a little bit of "value-added" and then sold for a big markup.

(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 think this will vary by fields, and I'm not qualified to speak in detail about any real fields - altho in the field I'm mainly interested, medicine, I would point again to numbers like numbers of individuals in a trial and compliance with various evidence-based medicine guidelines like blinding, randomization, intent-to-treat, etc. There's a vast amount of historical data that needs to be structured, and the project could open this up to the public to edit while allowing employees to supervise and do their own work.

(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?

Not sure. My impression is that the large publishers get such a big slice of the pie that if you cut their slice, the rest of the groups won't feel so much pain. Perhaps I'm wrong about that.

(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.

(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?

I don't know about this. I believe that scientists and peer-reviewers should be well-compensated for the work they do - I do not think that science is free or that it should always be free. My impression is that the current system does not do this efficiently. Embargo periods encourage payment, which is OK, but the payment is to the publishers, who then inefficiently transfer it to the scientists.

I am not a scientist. I got a couple BAs (economics, philosophy) in 2008, and have since worked as an insurance regulator in Alaska. I edit Wikipedia which is partly what my interest in publications stems from. I've benefited enormously from public access to articles, although to be fair the mere fact that abstracts are available online is probably even more of a benefit.

Regards,
Ben Creasy