

January 2, 2012

Office of Science and Technology Policy (OSTP)
and
National Science and Technology Council (NSTC)
Sent via email to publicaccess@ostp.gov

RE: Request for Information: Public Access to Peer-Reviewed Scholarly
Publications Resulting From Federally Funded Research

The Kauffman Foundation, as the nation's largest foundation devoted to furthering entrepreneurship and innovation, has a strong interest in high-quality social science research relating to the conditions under which entrepreneurs are most likely to develop and succeed. As well, the Foundation is among the largest, if not the largest, private funders of economics-based research in the United States. Taking into account both our purpose and similar role in funding research, we are pleased to offer comments on the Office of Science and Technology Policy's (OSTP) and National Science and Technology Council's (NSTC) request for information related to "Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research."

Before we respond to the specific questions raised in the request, we saw some topics that were omitted from the call which seem as or more important to advancing scientific innovation. Even more important than simply providing open access to peer-reviewed publications is the release of the intermediate data and methodologies that stand behind the research. Much scientific research is no longer able to be replicated, or even completely understood, simply by reading the 'methods' section of a scientific article. Research is now made up of complicated procedures, combined datasets, and complex computational algorithms all used to create new knowledge. Mandating that federally funded research make the entire scientific process more open and freely available will ensure that scientific results can be replicated, new methods can be easily created, and new knowledge can be built on what we already know. In addition, by creating a mechanism for publicizing computational methodologies, knowledge that can be used for commercialization will be disseminated and the creation of new companies will be fostered.

While we recognize this discussion is very discipline-specific, we see opportunity for cross-disciplinary impact through the creation of an online federal repository for the data and methods underlying federally funded research. While the actual repository would be just of federally funded research, the metadata, procedures, and overall processes could become standards within the industry. This is an area of standardization that, to our knowledge, has not advanced as quickly as some other areas of scholarly research search-and-storage practices.

Beyond this overall comment, we have arranged our comments directly in response to the requested questions from the Task Force on Public Access to Scholarly Publications:

(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 publicly 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?

Unfortunately, embargoes as currently used are entirely backwards when it comes to scientific freedom. There is immediacy to new scientific knowledge, so new papers are more valuable. However, we currently operate under the assumption that the papers at the frontier of knowledge should be trapped behind paywalls (at least for a certain amount of time), rather than open to all.

So, how should we think about embargoes for science? There are two components in new scientific knowledge: the data and underlying computational methods behind the research, and the final research results embodied in the scientific paper. While the data and computational methods are vital to a scientist's ability for new discoveries, the scientific paper is the end result and should be promulgated as quickly as possible.

Therefore, while an embargo can be constructed for the data and underlying methods, with public dissemination six to twelve months after publication through placement in a public repository, no such embargo should exist for scientific publications. Many newspapers allow free access to new articles, with articles older than a certain amount of time behind a paywall. A similar model could be used for federally funded research. During the first month of an article's release, when it is perhaps most scientifically valuable, an article should be open access. After this month, for example, a small charge could be levied for access to older articles. Of course, this shouldn't be a burdensome cost to the reader, as, ultimately, it is vital that federally funded research be open in order to encourage new knowledge creation.

But how could these be achieved? Ultimately, we want to focus on two steps that would encourage this possible end goal. First, the federal government must actively continue the person-level databases necessary to track the people undertaking federally funded research and the research they are undertaking. We will return more to that topic shortly. Second, federal funding agencies need to consider how such policies to

encourage open access to timely scholarship could be implemented without too much hardship on scholarly publishers. One possible solution would be a nominal open access fee that could become standard within federally funded grants above a certain threshold, much like administrative costs are standard in academic grant proposals. Institutions would include this very marginal fee in new grant proposals to federal funding agencies; it would be collected by research institutions in an “open access fund,” and then paid to publishers of academic materials when a journal article was accepted in recognition that the journal article would be made free to the public for a set amount of time.

(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?

How can journals be viable under this model? The traditional scientific publishing model is obsolete. It is time for the decoupled journal, where distinct functions of the journal—peer review, editing, and more—can be accomplished separately and more cost-effectively. Others, including Priem and Hemminger (http://www.frontiersin.org/computational_neuroscience/abstract/14455),¹ have written on this topic.

Ultimately, what is important is that we ensure that scientists are given credit for their own work and are provided a certain amount of priority in dealing with their own novel datasets and findings, and that science is disseminated rapidly so it can be built upon to create discoveries and even the commercialization of knowledge. Everything else—such as maintaining the current publishing business models—is secondary.

That said, we recognize that these academic systems are very rigid and that the inherent tensions in academia to publish or perish do not lend themselves to bold movements. If a means of compensating existing publishers for the important service they provide through incorporating something like the proposed “open access fee” into new grants made federally is not agreeable, then perhaps the federal funding agencies

¹ A pre-review copy is available at https://docs.google.com/document/d/1xDOy9GXXrUFc9TUIR2C470DTau8JEgZ9k-SMNlx5pb8/edit?hl=en_US&authkey=CMeCqOYD&pli=1. We recognize the irony that the final journal copy of the article is currently blocked for download by the general public; although we do not know if this research received any public funding.

and publishing industry could reach another agreement on how best to allow for open access using existing systems.

(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 commercial marketplace, from Google to many others, have actively expanded access to published research recently, even if some of the final articles eventually remain behind firewalls. If the federal government wants to encourage open access to its funded research, ultimately it has to be able to track the people producing that research much better than it currently does. While nascent efforts are underway through programs such as STAR Metrics, this work should be accelerated and, ideally, more information on the underlying names of people receiving federal funds made public. STAR Metrics uses administrative records from universities to collect person-level, longitudinal information on the people involved in carrying out federally funded research. Finding a way to expand STAR Metrics so that it begins to cover the universe of federally funded programs moving forward, integrate with other new efforts to create a uniform researcher ID (such as ORCID) that would track researchers across time, and ultimately use this data to interface with commercial publications databases is imperative. If existing databases (and their metadata standards) could be adapted to include a “federally funded” tag that was either designated by the author or arrived at through an administrative process, then this tag could be incorporated into the very smart systems that publishers have for tracking the places from which people are trying to download their publications.

It is of note here that we are primarily talking about a means of addressing this issue moving forward; addressing access for existing and past publications is an entirely different issue and one that probably would be best left until a later time. With timely research being the most important for driving innovation, putting in place a new regime of agreements and standards is most important.

(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?

There has been a proliferation of models in recent years, so we won't be able to give a full account except for those in which we are involved. We engaged a private company, Social Science Research Network (SSRN), some six years ago to help us expand the repositories available for entrepreneurship scholarship. Under the terms of our agreement, we provided seed funding for their expansion and a per-university flat fee for use of the network. Such a setup allowed SSRN to spend resources establishing the network; Kauffman money was spent only when users were actually using the network. SSRN is also an interesting example: By entering into hosting agreements with educational institutions around the world, it has worked around some issues about how its data ultimately will be curated. Other foundations, such as the Sloan Foundation, have interest in this space, too, so as public-private partnership is explored, perhaps we could help to be catalysts. Mellon's early involvement in setting EconLit is a great example of how we can help to influence this space.

(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?

This doesn't seem to be a big problem with final, published research, from our view. The march of companies like Google and others into this space seems to be pushing parties to have interoperable systems for the basics of a given article/citation. As stated previously, the problems lie more with the underlying data and procedures used to create the research. Those things are not currently very transparent.

(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?

We have proposed two actions that could lead to a possible solution: the "open access fee" and an expanded STAR Metrics (or similar) database. Our proposal has the advantage of working through existing infrastructures. The "open access fee" could be incorporated into new grant proposals to federal agencies, collected from the federal government by research institutions, and paid to journal publishers in return for open (or at least more open) access to the research when research is published. An expanded STAR Metrics program would create the people-level records necessary to more automatically track what federally funded research is being produced by whom.

Additional possibilities might exist around an expanded PubMed or other proven discipline-specific solutions.

(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?

The boundaries between forms of scholarship increasingly are blurred. We believe the focus should be on the peer-reviewed journal articles, realizing that this focus also will include many conference proceedings and book chapters. But the infrastructures for systematically addressing these latter two forms of scholarship in a standardized fashion are considerably more nascent.

(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?

As stated earlier, it's not clear to us that the focus of the embargo period is correctly placed. Priority should be given to open, immediate access with perhaps then the embargo period coming after a delay.

Please feel free to contact us (816.932.1000) to clarify any of these comments or request additional information.

Sincerely,



Samuel Arbesman, PhD
Senior Scholar, Research and Policy
sarbesman@kauffman.org



E.J. Reedy
Research Fellow
ereedy@kauffman.org