



Paul N. Courant

University Librarian and Dean of Libraries
Harold T. Shapiro Collegiate Professor of Public Policy
Arthur F. Thurnau Professor
Professor of Economics and of Information

818 Harlan Hatcher Graduate Library South
Ann Arbor, Michigan 48109-1205
734 764-9356 pnc@umich.edu

January 4, 2010

Dr. Diane DiEuliis
Assistant Director, Life Sciences
Office of Science and Technology Policy
Attn: Open Government
725 17th Street, NW
Washington, D.C. 20502

Re: Request for Public Comment on Public Access Policies for Science and Technology
Funding Agencies Across the Federal Government

Dear Dr. DiEuliis:

Thank you for the opportunity to comment on this important matter of public policy. President Obama's Memorandum on Transparency and Open Government of January 21, 2009 calls for an "unprecedented level of openness in government." When the nation pays to conduct research it seems unthinkable that the work is not already freely available to our citizens for the betterment of industry, education, and business. Information is expensive to produce and, in the current marketplace, expensive to share. In aggregate, libraries spend over a billion dollars each year to make information available — but even then access is limited by the scope of the licenses involved. Whether by subscription to journals or "pay per view" models, general use and browsing is prohibitively expensive for most people. This is a perplexing result for research that has been paid for by taxpayers. Most recently, the American Recovery and Reinvestment Act of 2009 appropriated \$17 billion to support research, research infrastructure and education, primarily through the National Science Foundation (NSF) and National Institutes of Health. How can the full impact of that research be felt if the research itself is not meaningfully accessible to our citizens?

Access to the best research done in the US benefits everyone and takes from no one. Thomas Jefferson's observation remains universal and true: "... no one possesses the less because everyone possesses the whole of it. He who receives an idea from me receives [it] without lessening [me], as he who lights his [candle] at mine receives light without darkening me." Today, the public does not readily have access to the fruits of much federally-funded research without paying explicitly for that access — despite having paid for performance of the work in the first place.

The University of Michigan Library has a long-standing commitment to the broadest possible access to scholarly materials. This is reflected in our efforts like Deep Blue (our institutional repository), HathiTrust (a national digital repository created by the cooperative efforts of some of

the nation's top research universities that is hosted by our library), or our Scholarly Publishing Office (which supports scholarship by providing sustainable electronic publishing services). These resources strive to provide open, web-based access to the research and scholarship where possible; access is limited only as required by contract or copyright. Unfortunately, these limitations are often binding. Members of the general public are free to search our catalog and holdings on-line, but they are generally not free to read the works unless our licenses with publishers permit public access. Thus it is a common occurrence that research funded by the US government, held in our library, can only be accessed by the University of Michigan community, on whose behalf we subscribe to scholarly publications.

We see our efforts here at the University of Michigan Library in the larger context of activities nationwide and worldwide. We are delighted but unsurprised at the success of the NIH mandate and PubMed Central as a free, publicly accessible, reliable source for NIH-funded research. It is a clear example of what is possible when the results of current scientific research can be widely and generally shared and repurposed. It would be a profoundly significant change to make the fruits of government-funded research openly accessible to the general public in a timely manner. It would also be relatively straightforward to achieve at little additional cost, taking advantage of existing infrastructures. Indeed, the cost of preservation, access, and dissemination of research articles would be modest as a portion of the overall cost of the research itself.

Please note that consistent with the Invitation to Comment, this response addresses specifically *scholarly publications* resulting from research conducted by employees of a federal agency or from research funded by a federal agency rather than other kinds of grant-funded activity (such as performing or fine arts), which may differ from scholarly publications both in type and modes of production. These are perhaps appropriate for another discussion in another forum, but I do not treat them here, as they are not scholarly publications as such.

My response to the questions raised in the Invitation to Comment follows.

1. How do authors, primary and secondary publishers, libraries, universities, and the federal government contribute to the development and dissemination of peer-reviewed papers arising from federal funds now, and how might this change under a public access policy?

There is a long-standing ecosystem of information in scholarly publishing. For decades, we have all participated in and necessarily relied on one another as authors, publishers, libraries, and universities to address aspects of peer review and distribution. Development of peer-reviewed paper is paid for by universities, research institutes, the contributed time of scholars, and, of course, granting agencies, notably including governments. Publishers, whether commercial or nonprofit, help to organize the review of scholarly work, and edit, "print," distribute and market the work. Selection, editing, distribution and sale are generally in the hands of publishers, a set of practices that made economic sense when printing and distribution were expensive.

But today's technologies allow for free and open access at essentially zero marginal cost. The issue is not and should not be "what has been" in the past or the preservation of no-longer-relevant business models for their own sake. The academy's publishing practices will change and are changing in response to new information technologies. The tie between publication and peer

review will almost certainly become weaker, as it already has in the fields that use extensive pre-publication sharing of work on the Web. But academic credentialing and peer review are properly the purview of the academy broadly, and their design and implementation should be orthogonal to the question of how best to make government funded scholarly research publicly available.

What might change under an open access policy? The Office of Science and Technology Policy's inquiry itself makes the case for public access to scholarly publications resulting from federally-funded research, listing several ways the return on federal investment in research may be better leveraged. To paraphrase some of the compelling reasons listed: the potential to promote advances in science and technology; improved cross-government coordination of government funding (and thus improved management of the federal research investments); more timely, easier, and less costly access to scholarly publications resulting from federally-funded research for educators and students, and "end users" of research, such as clinicians, patients, farmers, engineers, and practitioners in virtually all sectors of the economy. (*End user* is a rather clinical term in this case for our citizens and our entrepreneurs. We talk today of a knowledge economy. They (we) have a right to this information as part of that economy. Moreover, in the use of scholarship, there often is no end user; there is merely a next user. (Who is the end user of Euclidian geometry or differential calculus?))

No doubt there will be a number of changes in the distribution of scholarly work under a federally-mandated open access policy. Such changes, as always, will create both gainers and losers. It is likely that the commercial publishers will see a reduction in profit. It is likely that the business models used by many scholarly societies will have to be revised, and that universities and other supporters of research will want to work with societies to see how their vital work can best be supported. Provided that the system of scholarly communication continues to provide well-vetted work, and I am confident that it will, there is no general social loss associated with an open access mandate of the kind being considered here, and there is tremendous social gain in making the fruits of federally funded work broadly available.

2. What characteristics of a public access policy would best accommodate the needs and interests of authors, primary and secondary publishers, libraries, universities, the federal government, users of scientific literature, and the public?

The policy should require — mandate — open access, not merely request it. The experience of the NIH in implementing PubMed Central is informative.

The mandate should apply to the Version of Record. Certainly the mandate could require access to the final peer-reviewed version *sooner* than the Version of Record. However the final Version of Record itself should be made open after a fairly short embargo period – a year or two at most. The final peer-reviewed paper and the ultimate Version of Record may differ considerably, thus it is important that the Version of Record as the authoritative reference and scholarly artifact be made publicly available under a mandate. I discuss this further in item 6, below.

The policy should apply primarily to work published in peer-reviewed publications.

The policy should allow grantees to use grant funds, or to apply for supplementary funds, to pay the publication fees at open access or open access hybrid journals that charge fees.

The policy should let authors choose which open access repository to use, provided it meets certain conditions of open access, interoperability, and long-term preservation.

The policy should apply to articles that result from research funded in whole or in part by the funder's grant.

The mandate should be enforceable.

(Much of this list of features of a successful policy is based on a discussion of open access found at <http://www.earlham.edu/~peters/fos/newsletter/08-02-06.htm>.)

3. Who are the users of peer-reviewed publications arising from federal research? How do they access and use these papers now, and how might they if these papers were more accessible?

Today, users are limited to those who can afford to pay for access. Other than the articles deposited in PubMed Central, the fruit of federally-funded research is often distributed through private publishers that charge fees for access in one form or another. Typically, users are likely to be affiliated with universities or industry-specific companies that necessarily make paying for information a priority. Many articles may end up in discipline-based repositories that the general public is unlikely to be aware of.

Even when they can be found, articles often cannot be used. Currently, public indexing usually identifies bibliographic information — but not the article itself. Today, a quick check on a search engine will usually lead one to articles of interest; they are findable. But one will generally be unable to access the articles themselves unless one is affiliated with a university or other institution that pays for the privilege of access. Consistent descriptions alone will not make the articles themselves.

Would others use these papers if they were more accessible? What would people use them for? Presumably many people would use these papers if they were accessible. It is impossible to predict because we have no data upon which to base an estimate. What we know is that millions of potential users cannot access the material under current practices.

We do have some relevant anecdotal information. Our librarians tell us that when our students graduate and leave the University of Michigan with all of its rich resources, they are often surprised to discover that they no longer have access to our professional journals and databases. This is a rude shock for many who have become accustomed to easy access to a rich set of resources during the course of their studies. Our alumni ask us what they can do to obtain access, and we tell them that we cannot legally provide access to alumni without incurring large increases in license fees. The works upon which they rely are on the other side of a pay wall that we cannot afford to scale or breach. The lifelong learning that we know is crucial to a vibrant knowledge economy is impossible without access to the work done by federally funded researchers.

4. How best could federal agencies enhance public access to the peer-reviewed papers that arise from their research funds? What measures could agencies use to gauge whether there is increased return on federal investment gained by expanded access?

Simply having a mandate to make peer-reviewed papers that arise from taxpayer-funded research publicly available would be a tremendous first step toward enhanced public access.

In the scholarly community, one way to gauge increased return will be to look at citation patterns, which tend to show an increase when articles are made openly available — meaning freely available to anyone with access to the internet. We see that pattern in open access journals generally. It is too soon to see that impact specifically for citations from PubMed Central. The best current analogy relates to citations to or from open access articles in general.

Another concern is the integrity of and access to accurate data in the course of public debates. An example is highlighted by discussion on November 30, 2009 on the Dot Earth blog from *The New York Times*. The blog “examines efforts to balance human affairs with the planet’s limits.” A recent discussion about climate change and peer review addresses concerns about the integrity of data used in the science of global warming <http://dotearth.blogs.nytimes.com/2009/11/30/more-on-the-climate-files-and-climate-trends/>. The discussion focuses on questions about the availability and reliability of data sets and how science and related public policy affect each other accurately — or inaccurately — depending on the honesty of such data; whether such data is made available in its entirety — or not — affects the ability of others to make accurate analysis.

This question of the completeness and trustworthiness of data collected with public money is significant in and of itself. Yet there is another layer here. The discussion in the blog relies heavily on information in an article that was funded with the support of the National Science Foundation. The article includes a copyright notice for the year 2000 by the American Geophysical Union and may be obtained for a modest fee of \$9 (for non-subscribers) from the *Geophysical Research Letters* (“Causes of Global Temperature Changes During the 19th and 20th Centuries” (see <http://www.agu.org/pubs/crossref/2000/2000GL006109.shtml>)). The *Geophysical Research Letters*’ policy is to make articles open access for six months from first publication. After that, access is available by subscription or by fee.

In this case, the fee is modest, and I applaud the experimentation of the hybrid open access economic model of the journal. That said, should there be any fee for this article if it was funded by the NSF with taxpayer dollars? How would public policy and public debate about global warming be affected if the citizens had direct access to articles like this? Is the integrity of our public processes and democratic assumptions something that we can measure? What is trustworthy information?

5. What features does a public access policy need to have to ensure compliance?

A successful policy will make it easy for researchers and scholars to deposit articles, will mandate that deposit, and will make compliance with the mandate a quid pro quo for funding.

To elaborate on the issue of ease, partnering with trusted institutional partners — those with a proven record of providing broad access and digital preservation — to create workflows and systems will promote compliance on the part of researchers and ensure access for the public. The University of Michigan has, for example, worked with the NIH to act as a *publisher* in its deposit system, and in doing so helps to assure that our researchers comply with the NIH Public Access Policy (Division G, Title II, Section 218 of PL 110-161). Cultivating such partnerships and creating a distributed environment for preservation will result in a robust infrastructure that supports both researchers and the public.

With regards to the importance of mandates and the aforementioned quid pro quo, the NIH experience regarding deposit into PubMed Central is instructive. Per <http://www.nihms.nih.gov/stats/>, fewer than 650 articles per month were deposited before the mandate was required by law. Now that the mandate is legally required, the average rate of deposit in the most recent six months is over 5000 per month.

6. What version of the paper should be made public under a public access policy (e.g., the author's peer reviewed manuscript or the final published version)? What are the relative advantages and disadvantages to different versions of a scientific paper?

The public interest and our researchers are both best served by having the final published versions — the so-called Version of Record — available to them.

Under the present system, that interest is balanced against the research community's current reliance on publishers to manage some aspects of credentialing. While the researchers themselves perform both the research and the peer review, today most publishers handle the logistics of this review as well as performing additional copyediting functions. They pay for this editorial work via subscriptions, and the argument that this requires compensation and some say over how the results are distributed has merit, although it is less clear that a subscription model is necessary to pay for the requisite work.

In any case, exclusive control of distribution and access to the published work for the long duration of copyright is inconsistent with the public's right to have access to the research it funds (especially given that as a general proposition a term of copyright of "life of the author plus seventy years" could easily mean 150 years of copyright protection). For an interim period, as new models for distribution and credentialing are tested and developed, access to the final, peer-reviewed paper, cross-referenced to the published version, may be the best that we can provide. However, our ultimate goal must be deposit of and public access to the Version of Record, as soon after publication as is practicable.

7. At what point in time should peer-reviewed papers be made public via a public access policy relative to the date a publisher releases the final version? Are there empirical data to support an optimal length of time? Should the delay period be the same or vary for levels of access (e.g., final peer reviewed manuscript or final published article, access under fair use versus alternative license), for federal agencies and scientific disciplines?

The ideal is immediate public access, though publishers argue that such access might undermine their business model and stress current systems for managing a number of editorial functions they have traditionally provided. The optimal length of time to protect that business model varies by research discipline, but a period of 6-12 months from publication of the final version or Version of Record is an embargo that most agree is suitable to balance the competing needs of private enterprise and the public good. The benefit of association with established journals and the importance to the academic community of having immediate access are likely to be sufficient for publishers to continue to thrive (albeit with some reduction in profit in some cases) with a relatively short embargo period.

Regarding fair use, there can be no use without access, so this policy should focus on access to research. Fair use principles are essential to effective copyright law quite independent of the issue under discussion here. However, because whether a given use is legally a *fair use* is always determined on a case-by-case basis, it is not helpful in the context of ensuring broad public access. Alternative licenses have not evolved or converged on a standard as well established as current copyright law and are infrequently used; as such, they are also not yet useful in this context.

It is likely that the optimal embargo period varies by field, and I can imagine that the granting agencies are best situated to make such determinations in consultation with the research communities and with OSTP. That said, it may well be that the simplicity of a blanket rule (six months or a year) with an appeals process would outweigh the benefit from a more elaborate set of optimizations.

The implicit suggestion that the final peer-reviewed version be made open sooner than the Version of Record may provide a useful compromise. But the Version of Record should be made open after a fairly short embargo period — a year or two at most — even in this case.

8. How should peer-reviewed papers arising from federal investment be made publicly available? In what format should the data be submitted in order to make it easy to search, find, and retrieve and to make it easy for others to link to it? Are there existing digital standards for archiving and interoperability to maximize public benefit? How are these anticipated to change?

Today, universities, libraries, indexers, search engines, and publishers work together and have developed an infrastructure that addresses the question of how to make works discoverable. As long as the discovered work is also accessible, we can and do index and preserve and present it in ways that are flexible, addressing new formats and technical needs quickly — perhaps more quickly than any centralized body.

Decades, even centuries, ago a collaborative network of universities, through their libraries, made discovery and access to our collective knowledge its mission. We argue that as long as such a network has stewardship of this content — non-exclusive stewardship is appropriate here — finding and disseminating it will follow naturally. I think the following is instructive. The University of Michigan was an early partner in bringing scholarly journals online, starting in the pre-worldwide web days. During their reformatting and digitization projects, publishers relied on

our collections, and the collections of other research libraries, to fill in gaps in runs of their own journal back files. Their own archives were incomplete, in some cases missing years of content.

Archiving and interoperability standards exist. There are many to choose from, in fact, and converging on the appropriate ones will require coordinated effort, perhaps via a system akin to the Federal Depository Library system that served the nation's needs for so many years. Preservation and dissemination of the work is assured by such redundancy, and by limiting our reliance on organizations whose mission includes service to shareholders and who must be responsive to the demands of quarterly earnings statements.

9. Access demands not only availability, but also meaningful usability. How can the federal government make its collections of peer-reviewed papers more useful to the American public? By what metrics (e.g., number of articles or visitors) should the Federal government measure success of its public access collections? What are the best examples of usability in the private sector (both domestic and international)? And, what makes them exceptional? Should those who access papers be given the opportunity to comment or provide feedback?

With regards to a commenting function, there are two important cases to consider — commenting on policy proposed as a result of the material published in peer-reviewed works, and commenting on the peer-reviewed works themselves. The OSTP comment process, via blogs and public forums provides a model for how the former can occur. We should encourage such dialogue and continue to provide venues for it. With regard to the latter, here again a mechanism for feedback on the peer-reviewed work itself already exists. Peers communicate both informally via conferences, email, blogs, etc., and formally through published letters, clarifications, and other peer-reviewed papers. Preserving those mechanisms is also essential, as those mechanisms currently in place work quite well and probably would not be greatly enhanced by adding centralized forums or functionality to them.

Regardless, the first step towards making peer-reviewed papers more useful to and more often discussed by the American public is to make them available. The greatest barrier to use is, as Tim Berners-Lee has pointed out, the absence of content itself, not want of good interfaces. (See “The Next Web” presented at TED, February 4, 2009 http://www.ted.com/talks/tim_berniers_lee_on_the_next_web.html. Berners-Lee specifically talks about data, but the principle is the same.) Public and private sector solutions to usability problems abound, and will continue to flourish as universities, companies, and the federal government all continue their work in this area — some with an eye towards the public good, others with an eye towards market share and profitability. Quantitative measures for determining success abound as well — hits, downloads, citations, etc. — but we publicly fund research because we have, in the words of our Constitution, a national interest in “the Progress of Science and Useful Arts.”

Ronald Reagan famously invoked the image of a “shining city upon a hill” for our country, as did John F. Kennedy years before him. In terms of research, we have sought this ideal for the world. We now have the opportunity to assure that the doors to that city's library are always open, for the benefit of all. We will not and cannot know which visit to that library, which hit on a database, which download from a repository, or which subsequent published work will solve a

Dr. Diane DiEuliis
January 4, 2010
Page 9

pressing social, medical, or other problem, so we must first set our sights on accessibility, knowing that use and progress will follow.

Thank you for the opportunity to comment on this matter.

Sincerely,

A handwritten signature in cursive script that reads "Paul N. Courant". The signature is written in black ink and includes a stylized flourish at the end.

Paul N. Courant