

Testimony of

T. Scott Plutchak
Director, Lister Hill Library of the Health Sciences
University of Alabama at Birmingham

before the

Subcommittee on Investigations and Oversight

of the

House Committee on Science, Space and Technology

concerning

“Federally Funded Research: Examining Public Access and Scholarly Publication Interests”

March 29, 2012

Chairman Broun, Ranking Member Tonko, Members of the Subcommittee, thank you very much for the opportunity to speak to you today on the issues surrounding public access to the peer-reviewed reports of federally funded research. My name is Scott Plutchak, and I'm currently the director of the Lister Hill Library of the Health Sciences at the University of Alabama at Birmingham. The views I'm expressing today, however, are my own opinions and do not in any way represent any official position of my university.

Historians of the beginning of the print age refer to the fifty year period from about 1450 – when Gutenberg's press was invented – to about 1500 as the "incunabula period," from the Latin word meaning "from the cradle." This was a period of rich experimentation when entrepreneurs, scientists and scholars tested the opportunities offered by this powerful new technology. What resulted was the beginnings of the great age of print that we all grew up in.

We are now in the incunabula period of the digital age, and the opportunities before us are tremendous. Digital technologies for communicating scientific knowledge hold the promise for accelerating discovery in ways unimaginable to previous generations. We see them affecting every aspect of our social and economic life. Certainly the impact that they've had on my world of libraries and librarians has been earth-shattering, and equally so for the world of research and education that academic librarians have dedicated themselves to. We have it within our grasp to establish a new information infrastructure for the communication and advancement of scientific knowledge that builds on the best of what we've done in the past, while taking full advantage of the power of these new digital information technologies.

Unfortunately, the debates over access to the peer-reviewed journal literature that have taken place over the last decade or so have been unnecessarily contentious and have diverted energy and attention from what could have been, and should have been, a careful examination of facts and opportunities. The recent flurry of activity surrounding the introduction of the Research Works Act (RWA) and the reintroduction of the Federal Research Public Access Act (FRPAA) have resulted in a great deal of the kind of sloganeering, wishful thinking, and shading of fact in order to score rhetorical points that has characterized most of the public debate around Open Access during the past decade.

The most vocal of the Open Access advocates are lobbying hard for passage of the Federal Research Public Access Act, but I'm afraid they are being terribly short-sighted. Passage of FRPAA would be the digital equivalent of the first half of the 17th century, and we have the opportunity to do so much more.

The great achievement in scientific communication in the 17th century was embodied in the first two scientific journals – the *Journal des Sçavans*, first published in Paris in January, 1665, and *The Philosophical Transactions of the Royal Society*, appearing in London two months later. For the first time, scientific reports would be bundled together and disseminated systematically across Europe. It was the result of the best minds of the time using the latest technology to advance scientific progress.

We have the potential now for a similar significant leap forward, but the FRPAA approach risks actually take us backward – to a world in which individual reports of scientific results are isolated from context, held in scattered repositories, with very little in the way of connective tissue.

The true value of the peer-reviewed literature comes from context, when it is connected to the work that comes before it, that is laterally related to it, and that provides a foundation for what can be built upon it.

I completely agree with the notion that the peer-reviewed reports of federally funded research should be made freely available. In a digital world where distribution is cheap and easy, we certainly ought to be able to figure out a way to make this happen. But I want to see this done within a context that maximizes the value of those reports and that takes full advantage of the expertise that publishing professionals can bring. Mere access simply isn't enough.

The report from the Scholarly Publishing Roundtable,¹ which I was fortunate enough to be a member of, lays out the issues that need to be balanced in order to achieve a truly robust scholarly communication infrastructure that will take full advantage of the opportunities before us.

The key elements are these:

- A focus on the version of record. FRPAA, like the NIH public access policy on which it is based, settles for access to the author's final manuscript version, rather than the final published version – the version of record. In most cases, to meet a current need, this is probably sufficient. But a robust system of scientific communication that is expected to persist over time requires access to the final version, the stewarded version; the version that some entity is going to keep track of, ensuring that corrections are appropriately made, that retractions are handled when necessary, and that permanent access is ensured. At present, we are facing a world in which multiple versions of articles are available through various repositories and websites and it is increasingly difficult for the reader to determine which version they are accessing at any point in time. We need to be developing policies that minimize the potential for this sort of confusion, rather than exacerbating it.
- Standards for interoperability. Text-mining and data-mining published research reports enables deep analysis of multiple experiments, where individual results can be combined to provide accurate summaries across a broad range of experiments. When these sorts of meta-analyses are performed by humans, they can be extremely beneficial but they are very expensive and require a great deal of time. We know that there is tremendous knowledge embodied in existing research reports if only we could effectively analyze it. In order to facilitate this we need to emphasize a standardized structure for these reports. It is not necessary for them all to be in a single repository if they are structured in standardized ways. The National Library

of Medicine's "Journal Archiving and Interchange Tag Suite" is well on the way to becoming a de facto standard.

- A focus on preservation. Most of the scientific literature in biomedicine (the area that I know the best) is now "born digital," and, increasingly, there is no print counterpart. While print continues to be an important medium in some areas of the social sciences and humanities, this is rapidly declining. Although there are several promising approaches at play in providing permanent preservation to born digital documents, these are all still in the experimental stages. Any set of policies designed to provide access to the results of funded research must take into account methods for maintaining access for decades and even centuries into the future. This is not a trivial problem, and librarians and archivists understand that it is a critical one. A recent study reported that only 15% to 20% of the e-journal content in two major research libraries was being preserved by current initiatives.²
- A recognition that different disciplines may require different policies in order to achieve maximum benefit. Just as there is a broad range of adoption of "born-digital" objects across scientific disciplines, so are there differences in funding streams and in use of the literature. One of the most contentious areas in the public access debates has to do with the length of the embargo. The NIH policy specifies a maximum of 12 months; FRPAA would shorten that to 6 months. And yet it became clear to us on the Roundtable, as we investigated the differences among disciplines, the 6 months would be longer than necessary in some disciplines and too short to be practical in others. This highlights the fact that disciplinary differences are so wide that any single simple access policy that is intended to apply to all federally funded research will have negative unintended consequences in some areas. While policy development across funding agencies must be carefully coordinated in order to maximize interoperability, some flexibility must be allowed and even encouraged so that agencies can develop policies that are acutely tuned to the needs of their disciplinary communities.
- And yes, free access. Librarians, researchers, educators and publishers are all committed to achieving the widest possible distribution of the results of scholarly research. The opportunity that we have in the digital world is to achieve much wider distribution than could be imagined in the print world. The challenge is to do that in ways that balance the elements that I mention above. My experience with the Roundtable, and my other work with the librarian and publishing communities over the years, leaves me convinced that we will be more effective in developing those solutions when researchers, educators, publishers, librarians and the public work together than if we continue on a path that sets us at odds with each other. Across the publishing industry, in both commercial and not-for-profit sectors, experiments in open access publishing are proliferating. It should be clear to any objective observer that publishers are in no way opposed to open access. They are, quite sensibly, seeking to develop business models that keep their organizations healthy, while maximizing access to what they publish.

Although many in the OA community disparage the contributions of traditional publishers, it is noteworthy that both the NIH policy and FRPAA are explicitly based on the assumption that there is *something* that publishers are doing that is absolutely vital. Investigators are already required by their granting agencies to report the results of their research back to the agency and yet compliance with these requirements is shockingly low. A recent study estimates that only 22% of registered clinical trials whose results were mandated to be reported within twelve months of the conclusion of the study had done so.³ One could argue that better enforcement of the reporting requirements already in place at the funding agencies, along with mechanisms to make those reports easily searchable and available would do much more toward “making the results of federally funded research available to the public” than would result from the passage of something like FRPAA.

But it is clear that what OA advocates want is not just agency mandated reports – it’s the peer-reviewed papers that we crave. And despite the endless claims that since peer reviewers do not charge for their work, there can’t really be very much expense involved in doing peer-review, neither NIH nor any other agency has shown any inclination to set up their own peer review system. One might ask the question, if publishers add so little value, and peer review is accomplished at practically no cost, why not empower the agencies to set up their own peer review panels and cut the publishers out altogether? But however much derision and disdain the publishing industry has to put up with, it is clear that they continue to provide *something* that even the most fevered OA advocates believe is essential. The question we should be asking is how do we maximize that value. And yet, a study by Ross and others shows that the results of fewer than 50% of completed clinical trials even show up in the peer-reviewed literature within 30 months of the completion of the trial.⁴ Without tackling these issues, far more research results will still remain unavailable to the public than will become accessible.

The Research Works Act was a terribly mistimed rearguard action that was designed to forestall federal legislation of any kind. It is noteworthy that although the Association of American Publishers lauded the act, many individual publishers were quick to distance themselves from it. But again, it should be noted that the goal of the Research Works Act was not to combat Open Access – it was to forestall federal regulation. Publishers don’t object to open access, they object to federal regulation. This is hardly surprising. Businesses in general prefer less regulation to more regulation. There is nothing peculiar to the publishing industry in this.

Nonetheless, it seems to me that some level of government regulation is appropriate for at least that portion of the publishing industry that is focused on communicating the results of research. The investments made by taxpayers and the potential benefits to be gained from a well-organized and robust system of scientific communication are such that the inconveniences of some level of regulation are warranted.

But framing the question as, “How do we get better access to the peer-reviewed reports of research” is merely a pale version of the question that we *ought* to be asking, which is, “How do we take advantage of digital technologies to develop a robust and innovative scientific communication infrastructure that fully takes advantage of the potential of 21st century digital technologies.”

We have reached a point, after over a decade of squabbling about Open Access, when the terms of the debate seem to have come entirely down to whether or not someone supports something like FRPAA or something like the RWA. And yet, my experiences with the Roundtable, and the other work that I’ve done in this area over the past dozen years have convinced me that we can gain so much more than FRPAA offers. But to achieve this we will need to come together as a community that includes scholars, educators, librarians, publishers, and the public at large. We’re going to have to listen to each other and be bold and creative. We have to recognize that mere “access” without paying attention to the other aspects of a scholarly communication system doesn’t get us very much.

If something like FRPAA is passed into law, it will represent a huge missed opportunity. We have already wasted years on a battle that needn’t have been fought. As authorized by the America COMPETES Act in December of 2010, the White House Office of Science & Technology Policy (OSTP) is currently reviewing the results from its latest Request for Information. Their approach to these questions suggests that they are attempting to achieve the kind of balance recommended in the Roundtable report. Anything that the Congress can do to encourage policy development along those lines will be welcome. Anything that impedes or interferes with that work, such as the passage of FRPAA-type legislation will, paradoxically, make the goal of a truly robust open access infrastructure for scientific communication even more difficult to achieve.

Developing federal policies that will maximize the public’s investment in research and provide incentives for the development of a robust scholarly communication system is complicated, and achieving the appropriate balance of interests may not be as emotionally satisfying as advocating the simplicity of something like FRPAA. But the American public deserves to have us do this right.

¹ Report and Recommendations From the Scholarly Publishing Roundtable. January 2010. http://www.aau.edu/policy/scholarly_publishing_roundtable.aspx (Accessed 3/26/12).

² Preservation Status of e-Resources: A Potential Crisis in Electronic Journal Preservation. Fall 2011. <http://www.cni.org/topics/digital-preservation/preservation-status-of-eresources/> (Accessed 3/26/12)

³ Prayle AP, et. al. Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study. *BMJ* 2012;344:d7373

⁴ Ross JS, et. al. Publication of NIH funded trials registered in ClinicalTrials.gov: cross sectional analysis. *BMJ* 2012;344:d7292