



The Association for Research in Vision and Ophthalmology
1801 Rockville Pike, Suite 400 ■ Rockville, Maryland 20852-5622
arvo@arvo.org ■ +1.240.221.2900 (Tel) ■ +1.240.221.0370 (Fax)
www.arvo.org

January 12, 2012

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

On behalf of The Association for Research in Vision and Ophthalmology (ARVO), I submit the following comments in response to the RFI issued on November 3, 2011. ARVO is the largest and most respected eye and vision research organization in the world. Our members include more than 12,600 researchers from over 80 countries. ARVO encourages and assists research, training, publication and knowledge-sharing in vision and ophthalmology. ARVO publishes two medical/scientific research journals which are published online only and are hosted at HighWire Press which is considered by libraries internationally as a trusted site and archive. In mid-2012 ARVO will launch a new online-only journal on the topic of translational ophthalmic science & technology, which will also be hosted at a trusted site. In addition, ARVO voluntarily deposits complete articles of all NIH-funded research published in its journals in PubMed Central on behalf of authors and at no charge to the authors.

ARVO supports the principle of providing the public with access to the federally funded scientific research. However, we believe that releasing the peer-reviewed research articles in direct competition with scholarly publishers undermines the ability of associations and societies to maintain the high quality standards of selection, review, production, and publication as well as protection of the scientific record.

Scholarly publishers provide essential services that ensure the quality and integrity of journal content. Through peer review publishers and the scientific community identify scientific shortcomings and inadequacies which continue through the revision and re-review of articles. Over 50% as for some journals as much as 75% of submitted articles are ultimately rejected because of these inadequacies. The continuous feedback to authors through review and editing immeasurably improves the final published product. Publishers also serve as guardians of scientific ethics and standards to ensure accuracy, reliability, ethical treatment of patients and humane treatment of animal subjects.

In addition, in our opinion, the current NIH policy confuses the community and the public regarding the completeness of the “public” record and who the actual publisher of the scientific material is. NIH has established itself in direct competition with private publishers while using public taxpayers’ funds to complete their redundant work. These activities jeopardize the financial viability of journals, particularly those published by learned societies

and associations that are dependent on subscription revenue and author charges to sustain their journals and educational activities.

In response to the specific questions listed in the RFI, please accept the following comments.

- (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?**

RESPONSE: ARVO supports the practice of access to all scientific research to scientists, practitioners, and the learned public, not just access to federally funded research. Like most learned societies and scientific associations ARVO provides access through subscriptions to small and large institutional, commercial, and academic libraries worldwide and provides free and open access to the public within six (6) months of publication. Also like most scholarly publishers ARVO's journals are hosted online in internationally recognized sites that are considered by the library and academic communities as "trusted" sites. ARVO's journals are hosted currently at HighWire Press. As with most non-profit, scientific associations, we are committed to the preservation and archiving of all of our content through participation in programs such as LOCKSS, CLOCKSS, and Portico, through which participating institutions may download and store as a deep archive all content from Volume 1, Issue 1, page 1 to current data. The deep archive can be used to restore data lost at the institution and to update files to meet then current technological standards. ARVO, through its electronic host, archives XML files and all metadata, as well as figures and tables in their native format so that the content can be repurposed or configured to meet future technological needs and formats. Internal links in articles to already published content and well tagged metadata allow robust search engine results that ensure discoverability and, thus, increase productivity for those engaged in the scientific enterprise. For Federal agencies to create additional archives or access points appears to be a redundant and inefficient use of federal funds that could diminish funds available for ongoing scientific research.

For example, the NIH requires deposit of all federally funded, peer-reviewed scientific journal articles. It then reprocesses all files, using at least one **non-US vendor** to do so. This use of U.S. taxpayer funds to support non-US vendors does not meet the stated criteria. The development of these materials housed at PubMed Central is redundant to the online content hosted and archived at trusted resources.

- (2) What specific steps can be taken to protect the intellectual property interests of publishers, scientist, 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?**

RESPONSE: Having a single source of the definitive scientific content of a journal ensures the integrity and completeness of the scientific record; that source should be the original article published by in journal. For most journals that have an electronic publishing component the official article of record is the electronic, XML/HTML version of the article. For example, at trusted sites if an erratum (correction) or comments about a specific article are published they are linked to the original article in perpetuity. With multiple sites the likelihood of constructing and maintaining the complete record with correct links is more problematic and could even endanger the lives of patients if the record is incomplete. One of the hallmarks of scientific research is the ability to duplicate results; if future research determines a flaw in the original findings, it is of paramount importance that the original research be linked through references and use of DOIs to future findings. These links (called forward linking) are maintained by most scholarly publishers today. The use of embedded DOIs (digital object identifiers which are discrete for each article and include a publisher's identification) for all parts of articles, including figures, tables, and text units, would also ensure that intellectual property rights are maintained. DOIs are searchable and are associated with the original **publication of record**—the journal article and thus the authors and publisher. We suggest that policies encouraging the use and deposit of DOIs by all publishers for all articles be considered.

Conversely, establishing multiple deposit sites with varying times of free and open access to articles/content is counterproductive from a scientific integrity, content management, and financial standpoint. Publishers have already made the investment in time, scientific resources, and financial resources; multiple other site hosts is redundant and financially wasteful. In addition, sites such as PubMed Central (under the National Library of Medicine (NLM)) have on part of the body of knowledge in any given medical/scientific field—that content funded by a Federal agency, specifically the National Institutes of Health. The NLM also is selective in which journals are included in their indices. It usually takes two to three years after a medical journal begins publication for the NLM to consider adding it to the collection. In the case of new journals that would mean that up to three years of content would never be included in the Library.

- (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?**

RESPONSE: Given the technology that exists today there is one rationale for a centralized repository and that is the development of specific content databases and analytic tools that aid in data mining. However, the shortcoming to that rationale is that if the database only contains federally funded content then potentially greater than 50% of the body of knowledge in any given scientific area would not be included in the database. Is the government in a position to **at least double** the expense currently incurred to development and maintain the content? The private sector has the incentive and expertise to develop new tools to meet the needs of all stakeholders, including search engines that can identify all content that is appropriate for inclusion, including across disciplines. Given the nature of translational

research, it is highly likely that relevant content in a particular area may be published across journals, and not necessarily in journals included in NLM or PubMed Central and possibly not funded by a Federal agency, such as new developments in engineering fields that directly affect medical technologies. Again, leading to an incomplete picture in the government's centralized repository.

If a Federal agency maintains custody of all published content and controls access to that content, non-profit associations and societies would not be able to sustain a publishing program and several thousand journals would cease publication. The remaining commercial publishers could not and, we believe, would not be able or willing to absorb the content. If non-profit organizations lost their publications and the associated revenue it would significantly and negatively impact the educational and support services that organizations provide to their members. Journal revenue helps support all other facets of non-profit organizations, including scientific meeting large and small. In ARVO's case, at least three small, highly specialized meeting would be cancelled; the meetings lose money and are almost entirely supported by publication revenue.

The government can ensure long-term stewardship of decentralized content by working with publishers and organizations such as NISO to quickly develop standards that must be met. If hosting sites were evaluated in terms of these standards as well as for the technological reliability and discoverability of all content, then the sites could be deemed "trusted". A regular monitoring of and support of new host sites as well as established "trusted" sites could ensure compliance.

- (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?**

RESPONSE: There are three well-respected models of existing publisher-library-archive partnerships that exist and that archive ALL content, not just federally funded research: Portico (portico.org); JSTOR (jstor.org); and ITHAKA (ithaka.org). Portico has 135 participating publishers with over 12,300 individual titles and over 19.35M individual items preserved. Other groups that permit local library archiving of titles to which they subscribe include LOCKSS and CLOCKSS. Portico and JSTOR have robust tools and are committed to full archiving of all content as well as providing access to content if a publisher ceases to exist or ceases publication. A private company that allows subscriptions to content as well as indices is Thomson/Reuters Web of Knowledge. This is a very large database that is not always current and is very rigid in its requirements. It does not use the DOI despite repeated requests from large and small publishers.

- (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-review publications resulting from federally funded scientific research are publically available to ensure that these publications can be easily found and linked to Federal science funding?**

RESPONSE: The minimum metadata that needs to be made available is the full citation of an article. That is, the authors names, title of article, journal name/abbreviation, volume, issue, page, year. This would allow searching down to the article level, if needed, or at a journal level. It would be useful for all publishers to incorporate the article DOI into the article citation. Using any search engine the online article would be retrievable regardless of file location. This is standard metadata for all scientific journals, most of which now use the NLM DTD (a DTD is a document that defines how all elements in an article should be identified (tagged). The National Library of Medicine is the most commonly used DTD worldwide.) The metadata is currently available through the National Library of Medicine for most medical journals. However, please note, if someone uses a general term in a search engine, such as glaucoma, articles including that word will appear. Scientists, clinicians, and researchers are very sophisticated technology users and have little difficulty identifying relevant data across platforms, publications, or archives. Journals include funding information, Federal or otherwise, in the body of their articles and do include the citations of articles published (with links if required) in all grant and support applications.

(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?

RESPONSE: The best way to maximize the benefit to all stakeholders is to allow publishers to set reasonable times to full open and free access to their content based on the frequency of publication (monthly, bi-monthly, bi-weekly, etc.) and the timeliness of the content. In this way publishers can control their expenses and maximize the subscription revenue that results in 30-50% of revenue of most scholarly journals. Library organizations have stated that if a journal is published monthly, most librarians would not consider dropping the subscription if the journal became open and free at 6 months or longer after publication. That is, if a January issue is published January 1, and that issue became open and free on July 1 of the same year.

Since publication costs (charges to authors) are generally low (less than \$3,500) relative to the size of grants, these costs should be accommodated in all grants, with no expiration date of funds being paid to publishers. A common response to receipt of an invoice for publication charges is “my grant has expired and I don’t have the funds.”

Again, it should be mentioned that establishing government archives for federally funded articles is redundant and expensive. Publishers are already providing archives and are paying for them. Realistically, the average American citizen is not reading scientific journals and most do not have the training or experience to evaluate the content for relevance. Patient advocates claim that everyone should read the articles or be able to access them. In fact, most medical publishers will provide single copies of articles to patients or family members of patients with the recommendation that **they take the article to their treating physician to determine the relevance and to aid in asking questions.** These articles are not intended for lay people; they are written by and for experts in their fields.

- (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?**

RESPONSE: ARVO does not believe that book chapters should be covered by these policies. Publishing textbooks or scholarly tomes is as expensive, if not more so, as publishing a journal. If chapters will be given away publishers may stop publishing any books that contain federally funded material. Regarding meeting abstracts, ARVO's meeting abstracts are free from the day they are launched. As abstracts are usually partial or preliminary findings, they should not be considered the definitive information on the topic.

- (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 weight public and private benefits and account for external and 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?**

RESPONSE: This is an extremely difficult question to address. So much depends on the timeliness of the material, that is how important is the content to the community today, in 6 months, in 12 months, so that most researchers/scientists would have read the material by that date. ARVO has determined that the highest "reader" rate is achieved in the first 6 months of publication. While the readership has a long tail, that is, it will be read regularly for about 6 years, and then sporadically after 6 years but will continue to be read when it is 10 – 15 years old. ARVO spoke with subscribing libraries regarding its largest journal, which is the leading research journal in its field in the world, and learned that generally they would not drop the subscription even if the articles were not free and open for 12 months, and then ARVO decided that as a service to the community, to help stimulate research, and provide timely content, particularly to researchers in developing countries, it would open all articles to public access after 6 months in its largest journal. ARVO's second journal is open access immediately upon publication and has been since its launch 10 years ago. However, to sustain this model of open access there are author charges that must continue to increase in order to meet the costs of producing and hosting the journal online. ARVO is, therefore, comfortable with a 6-month embargo period.

Thank you for the opportunity to respond to this import RFI.

Submitted by:

Karen Schools Colson

ARVO, Director, Publishing Projects

On behalf of The Association for Research in Vision and Ophthalmology

1801 Rockville Pike

Rockville, MD 20852