



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 Digital Data Resulting From Federally Funded Scientific 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, publication and 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 continues 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.

ARVO appreciates your consideration of our responses below to the specific questions posed in the RFI regarding *Preservation, Discoverability, and Access and Standards for Interoperability, Re-Use and Re-Purposing*.

- (1) **What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?**

RESPONSE: Since most association and non-profit publishers in all scientific and medical areas already offer free and open access to ALL content, whether or not it is federally funded, anywhere from three (3) months to twelve (12) months after publication and usually provide open access to the abstracts immediately upon publication, ARVO does not believe that additional policies would improve the access or dissemination of information. In addition, by requiring federally funded research to be available earlier only presents 30% to 50% of the published research, thus presenting an incomplete resource on any given topic. Increasing funding for scientific research would better serve the public, the economy, and researchers. Like most association and non-profit publishers, especially those hosted at trusted sites, ARVO also participates in the LOCKSS (Lots Of Copies Keep Stuff Safe) which permits any interested libraries to deep archive all journal content and update those files for future use if deemed appropriate. It would better serve the public to cease requiring deposits of published articles in a taxpayer funded site such as PubMed Central (PMC), which is redundant and by all reports from the National Library of Medicine costs over \$2,500 per article to create and maintain, and allows indiscriminant distribution of all content.

It should be noted that private sector publishers represent over 30,000 workers in the U.S.; spend millions of dollars to provide peer-review, editorial support, and production and distribution of over 45% of the scientific peer-reviewed articles published each year for researchers around the world.

- (2) **What specific steps can be taken to protect the intellectual property interest of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?**

RESPONSE: As mentioned in Response to Question (1), publishers should be permitted to control the access and distribution of their content and are currently providing wide access within one year of publication, which often includes all back issues, as is the case with ARVO. We provide free and open access to all content published from Volume 1, Issue 1, page 1 of our journals through all information published up to 6 months ago. We are investigating the possibility of watermarking all content to ensure that the copyrighted content cannot be indiscriminately copied, reused, and redistributed without providing appropriate credit and attribution to the authors and publisher. This would apply to all supplementary data that is published with an article. Our intent is to protect the integrity

of the research and ensure proper accreditation of the research. In addition, agencies should consider supporting America's Research Act, H.R. 3699.

- (3) **How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?**

RESPONSE: Regarding online only publications in the medical area, PubMed Central has recently advised published that ALL content, regardless of whether it is federally funded, must be deposited and made freely available when ready through PMC. This, in our opinion is counter to the Digital Millennium Copyright Act and US Copyright Act, Title 17. It seems reasonable that Federal or federally trusted, perpetually maintained data repositories could be established to store data tables or genetic/genome information databases in which all authors could deposit such content. Medical and scientific associations and non-profit publishers are already providing trusted sites and archives for all of their content and these facts should be considered before requiring additional federal resources be used to establish a redundant system of managing data. In these days of world-wide discoverability through the use of robust search engines, associations and non-profit publishers are already meeting data management and discoverability needs.

- (4) **How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?**

RESPONSE: ARVO suggests that as a first step, agencies should evaluate **existing** trusted source resources that currently exist and examine the practices for access, dissemination, and preservation and that are currently being paid for by publishers. These sites would include HighWire Press, Allen Press online, and such organizations as the American Geophysical Union. Establishing any federal repository that would include data already stored in such a site would be redundant and an unnecessary use of taxpayer funds that could be used for supporting additional research.

- (5) **How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?**

RESPONSE: If stakeholders participated in already existing plans, such as LOCKSS, CLOCKS, and other deep archives, as well as agree on requirements for qualification of trusted hosting sites then all could begin public discussions of data management standards.

- (6) **How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?**

RESPONSE: First, and foremost, an evaluation of the costs incurred in maintaining already established archives and deep archives should be undertaken. ARVO suggests that organizations such as HighWire Press, The Stanford Libraries, and OCLC's activities, as well as others such as Portico, should be reviewed and considered. These organizations have already committed to digital preservation and perpetual access. Their costs could be used as benchmarks for current and projected costs to best address funding levels and mechanisms.

- (7) **What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?**

RESPONSE: If a centralized list of trusted hosting and perpetual archiving resources were developed and associated with recognized scientific journals and publications (books, meetings abstracts, etc.) then authors could report where their research is published and be in compliance with any and all funding institutions. This would minimize cost of compliance and verification. Since publications lists are an integral part of grant applications this would further consistent reporting of compliance.

- (8) **What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?**

RESPONSE: Encouraging the continued development of affordable semantic search and discovery tools and supporting the publishers in the use of these tools would provide stimulus to a variety of markets and industries. It is important to support innovative U.S. companies in these any other technology areas.

- (9) **What mechanisms could be developed to assure that those who produce the data are given appropriate attribution and credit when secondary results are reported?**

RESPONSE: One method would be to support the use of Digital Object Identifiers (DOIs) for all elements and parts of all published content and embedding of DOIs in figures, tables, and content and parts of content. Publishers already use DOIs and deposit the information with groups such as CrossRef. Whenever a DOI is searched or used it is associated with the original citation. If DOIs were used universally by indexers and abstracters and checked verified by publishers during the publishing process and use a reverse lookup at regular intervals after publication, original publication information of results would be readily apparent and available. Note that over 1,200 publishers, libraries, and organizations world-wide are members of CrossRef and already support the use of DOIs. The organizations also agree on standards for use and presentation of DOIs for all facets of scholarly publishing. CrossRef also participates in establishing NISO standards for use.

Standards for Interoperability, Re-Use and Re-Purposing

- (10) **What digital standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, Nature Genetics 29,371) is an example of a community-driven data standards effort.**

RESPONSE: ARVO has no definite response to this question at this time other than allowing each scientific community to develop its own standards.

- (11) **What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?**

RESPONSE: ARVO has no definite response to this question at this time but agrees with the need for standards but recognizes the need for flexibility in the standards because of ever-changing technologies and discoveries.

(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?

RESPONSE: One method of promoting effective coordination internationally is to work with and through international specialty organizations such as ARVO. Many associations and scientific societies, including ARVO, have members worldwide who represent government agencies and institutions in their home countries. This would be a very effective means of developing good lines of cooperation and communication. Our organizations continually promote and support international collaboration in research.

(13) What policies, practices, and standards are needed to support linking between publications and associated data?

RESPONSE: Please refer to earlier discussion of DOIs in Question #9.

Thank you for the opportunity to respond to this RFI issued by OSTP.

Submitted by: Karen Schools Colson

ARVO, Director, Publishing Projects

On behalf of The Association for Research in Vision and Ophthalmology

1801 Rockville Pike, Suite 400

Rockville, MD 20852