ASSOCIATION OF RESEARCH LIBRARIES

**Comments of the Association of Research Libraries Concerning
"Public Access Policies for Science and Technology Funding Agencies
across the Federal Government"**
Submitted to the Office of Science and Technology Policy, January 15, 2010

Summary

Thank you for the opportunity to comment on "Public Access Policies for Science
and Technology Funding Agencies Across the Federal Government." Enhancing
public access to federally funded research results has been and continues to be a
priority for the Association of Research Libraries (ARL) and its member libraries.
We very much appreciate the interest and focus of the Office of Science and
Technology Policy in stimulating a public discussion regarding the benefits of
enhancing public access to federally funded research. ARL supports enhanced
access to federally funded research resources because such policies are integrally
tied to and support the mission of higher education and scholarship. ARL
believes that extending public access policies to federally funded research to
other science and technology agencies will be a central component of President
Obama's transparency and open Government initiative. We fully support such
an extension.

ARL is an association of 124 research libraries in North America. These libraries
directly serve 4.2 million students and faculty and spend $1.3 billion annually on
information resources of which 45% (in 2008) is spent of electronic resources.

New investments in cyber and information infrastructure are critical components
to advancing science and education and spurring innovation. Reports such as
the National Academies report, *Rising Above the Gathering Storm: Energizing and
Employing America for a Brighter Economic Future*, reflect the pressing need to
ensure an environment that is conducive to enabling the United States to meet
the global challenges of the 21st century. This means that researchers, students,
and the public must be empowered by having the full array of information
resources needed to make contributions in all arenas. Broad distribution of
information and research enables scientists, including citizen scientists and
university researchers, to build upon it and approach problems with new
perspectives. It permits educators and students to have access to needed
resources previously unavailable, without regard for geographic location or
financial limitations. And it gives members of the public access to resources that
they have paid for and may require in their daily lives or in support of
educational interests.

Many discoveries result from building on prior studies. For example, the
discovery of the structure of DNA, the development of penicillin, and the
development of radiation treatment for cancer patients all stemmed from
researchers building on the work of others. It is time to extend discovery and
access well beyond current bounds to foster new educational applications and
advance science.

Through an Executive Order or working with congressional leaders on a legislative approach, the Obama Administration should mandate that all grantees who receive federal funds from an agency be required to deposit either the final, published version of a peer-reviewed journal article or the final electronic manuscript of such an article in a publicly available digital repository. The role of the digital archive would be to provide long-term curation and access to this literature and to be interoperable with other digital archives and institutions. There should be no restrictions placed on use of this literature, on who is able to use these federally funded information resources, and the embargo period, if there is one, should be as short as possible.

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?**

Federal agencies, universities, researchers and authors, libraries, and publishers contribute to the scholarly publishing enterprise in different ways and to varying degrees. Agencies fund researchers who increasingly, through collaboration with others nationally and internationally, conduct research, disseminate their work through multiple channels and engage in peer review on a voluntary basis. Universities play a key role in supporting their faculty, researchers, and graduate students in multiple ways, including the contribution of time to undertake peer review, funding of their laboratories and research libraries, and more. Publishers (both not-for-profit and commercial) help to organize the peer review, copy edit, publish, and disseminate the peer-reviewed works in print (decreasingly available) or via licenses electronically. Research libraries license these journals and in turn make them accessible to members of their campus. When possible, these libraries also preserve these works. More recently, research libraries have established institutional repositories comprised of the intellectual content of their institution and are working with agencies such as the National Institutes of Health (NIH) in support of public access policies. Although license terms and conditions prohibit many research libraries from making non-open access (OA) journal literature more broadly available, many public research institutions do provide access to these articles (licensed resources) to members of the public who physically come to the research library.

Faculty, researchers, and students affiliated with our institutions conduct and collaborate on research and share the results of their research in support of the scholarly and scientific enterprise. As noted in the AAAS report, *Intellectual Property Experiences in the United States Community*, the primary motivation for scientists to publish their works is "to inform others about their work." (http://sippi.aaas.org/Pubs/SIPPI_**US**_IP_Survey.pdf, page 8). Thus, providing greater access to these works through new and/or extended public access policies of federal science and technology agencies is completely consonant with scientific practice.

The widespread use of information technologies has fundamentally changed the conduct of science and is now changing how scientific research results are made available. Scientific and research communities conduct research, for example, though computational methods such as the mining of scientific literature and data. Publication of peer-reviewed research results needs to reflect and incorporate this evolving practice. This change in practice has led to new roles for research libraries. For example, most research libraries have established institutional repositories to collect, maintain, preserve, and provide access to the intellectual content of their institution. With the implementation of the NIH Public Access Policy, universities and their research libraries are providing new services, as this policy presents an opportunity for these institutions to support their faculty and researchers in new ways while meeting federal requirements. It is anticipated that these new campus-based roles will continue to evolve and that the number of institutions working collaboratively with federal agencies in this capacity will increase. It is exactly this type of innovation that enhanced public access policies should stimulate.

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?**

There are a number of key characteristics or components that frame successful, existing national and international public access policies. First, one integral component of a public access policy that can be undertaken prior to and after implementation is appropriate consultation with affected communities. Before implementing the NIH Public Access Policy, the agency engaged in extensive outreach to various affected constituencies. This consultation has continued, and the NIH Policy evolved to reflect these discussions. Even though NIH, like all federal agencies, has the authority to implement enhanced public access policies (e.g. federal purpose license), they chose to work with the various stakeholders to fashion a policy to meet community needs. Similarly, the Office of Science and Technology RFI reflects this type of public consultation. Consultation should occur within the normal legal parameters to elicit valued input; however, care should be taken to ensure that such consultation does not slow down or inhibit agencies from instituting public access policies.

Second, an important outcome of the NIH Public Access Policy discussion and early implementation was the change in policy from voluntary deposit of the final electronic manuscript to mandatory deposit. Any new public access policy relating to federally funded research should learn from the NIH experience and that of other countries and mandate deposit of the final, peer-reviewed electronic manuscript or published article.

Third, accountability and long-term access to federally funded research results are critically important and compelling reasons that governments invest in public access to federally funded research policies. Thus establishing stable, interoperable, and permanent digital archives are additional characteristics of

robust public access policies. Such archives do not need to be within federal agencies and can be found, for example, at universities or other institutions.

Fourth, national and international public access policies call for <u>inclusion of peer-reviewed literature</u>—either the <u>final, published version of the article or as, in the case of the NIH Policy, the final, electronic peer-reviewed manuscript</u>. The final, published version is clearly preferable, but only if there is permission of the copyright owner, and importantly, only if users are able to fully use and reuse the article without limitation. <u>Full use rights</u> (e.g., data mining, linking to and more) are essential components of successful public access to funded research policies. This is a key enabler of research and discovery. Such use rights are in contrast to only permitting the right to read and access a work. Utilizing the final, published version may entail some additional delay in access, so deposit in a public repository of the final electronic manuscript for use until the final, published version is available is an important consideration. Ideally, such an <u>embargo period should be as short as possible.</u> The common embargo period for public access to funded research policies is six months, with NIH's embargo period of 12 months being the maximum. Of course, if articles are published in an open access journal, there is no embargo period.

Finally, other characteristics of public access policies relating to funded research concern <u>compliance issues.</u> For example, grantees must retain certain rights (e.g., the ability to deposit in a public access repository) as a condition of funding in order to comply with the agency policy, and the agency must implement compliance mechanisms to ensure that the policy is successfully enforced.

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? Would others use these papers if they were more accessible, and for what purpose?**

Today, with the exception of federally funded research literature in open access journals, timely access to peer-reviewed publications is limited by subscription barriers and embargo periods. Research libraries, on behalf of their faculty, researchers, and students, are the predominant subscribers to scientific and research publications, and even these institutions are unable to afford all of the needed journal literature. Embargo periods of 12 months or more constitute too much of a delay for the research community; thus, research libraries must subscribe to these journals for their users.

As mentioned previously, licensing terms and conditions (to non-open access journals) restrict use of the journal articles. As a result, members of the public—citizen scientists, students, teachers, small business owners and others—must subscribe to journals, which in some cases can range from $630.00 for the New England Journal of Medicine to $29,914.00 for Brain Research, or purchase, when possible, individual articles. The cost of each article is not trivial (usually more than $30.00 per article within the science, medicine and technology arena), again limiting access by non-academic users.

As has been demonstrated with the implementation of the NIH Public Access Policy, the daily use of peer-reviewed literature in PubMed Central (PMC) is significant and growing.

Number of PMC Articles and Monthly Access
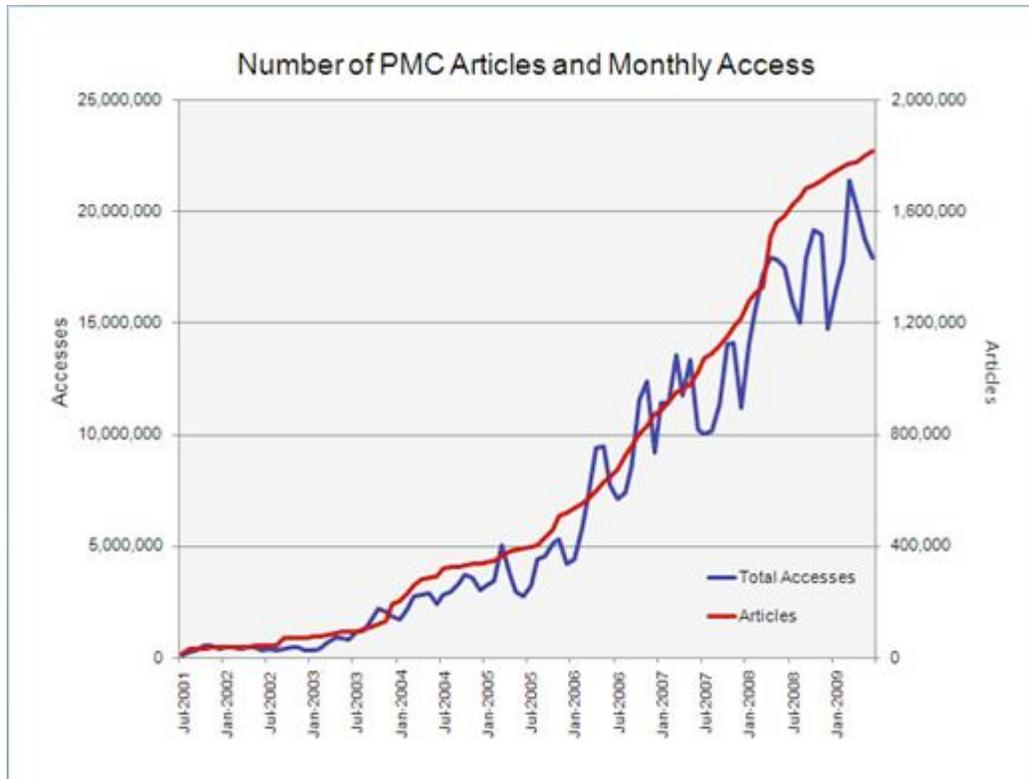
The ability to read, data and text mine, link to, and discover data and information should be available to all members of the public who have funded this research, especially as more and more individuals search the Internet for needed information. Importantly, making these research articles publicly available levels the playing field for many institutions that are unable to afford access to needed research materials.

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?**

Agencies could mandate that all grantees who receive federal funds from an agency be required to deposit either the final, published version of a peer-reviewed journal article or the final electronic manuscript of such an article in a publicly available digital repository. The role of the digital archive would be to provide long-term curation and access to this literature and to be interoperable with other digital archives, resources and institutions. There should be no restrictions placed on use of this literature or on who is able to use these federally

funded information resources, and the embargo period, if there is one, should be as short as possible.

The NIH Public Access Policy is one approach that agencies could consider. NIH via the National Library of Medicine (NLM) has a long history of collecting, maintaining, preserving, and providing access to biomedical literature. NIH has carefully tracked usage of PubMed Central, the costs of implementation of their public access policy, and of its benefits to NIH and beyond. As one of the largest funders of research in the Government ($30 billion in FY 2009 with additional stimulus funding), the number of articles arising from this funded research is approximately 80,000 per year. This number is likely far higher than that of other federal agencies; accordingly, it is possible that agencies will design public access policies that meet the unique needs of their mission and constituencies. In addition, NIH has written source code that is portable to other agencies. If agencies choose to utilize the PMC approach, the costs of implementation will be reduced. Overall, NIH spends several million dollars per year on ingesting 80,000 articles into PMC. This is a small amount compared to the agency's spending on research and to the value that is now available to the public. PMC's database is one part of a suite of valuable public resources that are accessed by more than 2 million users every day.

Other agencies can learn from NIH's experience and build on NIH's success, either through implementing a similar service or by designing a public access program that better meets the needs of that agency. For example, working with the Wellcome Trust and other United Kingdom research funders, UKPMC began operations in 2007. UKPMC shares journal content with PMC and the repositories are interoperable. Similarly, PubMed Central Canada is a joint project of the Canada Institute for Scientific and Technical Information (NRC-CISTI), the Canadian Institutes of Health Research (CIHR), and the US National Library of Medicine (NLM). Finally, in 2009, the National Science Foundation (NSF) awarded the Sheridan Libraries of Johns Hopkins University $20 million to build a data research infrastructure for the management of digital information created for teaching and research. Beginning with the life, earth, and social sciences, project members will develop a framework to more fully understand current data practices and develop a model for curation that allows ease of access both within and across disciplines. The Libraries also received $300,000 for a feasibility study of developing, operating, and sustaining an open access repository of articles from NSF-sponsored research. These examples provide direction for other agencies as they explore how best to implement new public access policies relating to federally funded research.

Whatever path agencies choose in implementing a public access policy, consistency of requirements is important. It will be difficult for research and academic institutions to comply with policies that contain different mandates and requirements. Ensuring relative consistency across agency policies is one key element to ensuring a valuable return of investment and foster a culture where sharing of these resources continues to promote the interests of science. Additional research and analysis can be found in the following studies.

- Houghton, J.W. Steele, C. and Sheehan, P.J. (2006). *Research Communication Costs in Australia, Emerging Opportunities and Benefits*, D*epartment of Education, Science and Training, Canberra.*

- *Department of Trad*e and Industry (DTI) (2007). *Measuring economic impacts of investment in the research base and innovation: a new framework for measurement*, Department of Trade and Industry, London.

- Houghton, J.W et al, (2009). *Economic implications of alternative scholarly publishing models: exploring the costs and benefits*. JISC EI-ASPM Project. A report to the Joint Information Systems Committee (JISC). London: JISC. (http://www.cfses.com/EIASPM)

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

There are several key features that should be included in a public access policy. First, as mentioned previously, experience has shown that <u>mandating deposit</u> of articles or final, peer-reviewed electronic manuscripts arising from federally funded research is critically important. Second, NIH and other governments engaged in <u>extensive education</u> with the grantee community in order to explain the Policy, its components, and its benefits. Third, <u>ease of compliance</u> has proven to be an important factor in the success of the various policies. For example, NIH has worked with publishers, libraries, and others in designing deposit systems that ease the burden on the individual researcher, the institution, and publishers. Finally, <u>consistency across federal agency policies</u> is important. In working with research offices, centers, and others on campus who are responsible with the grantee for compliance, ARL has heard repeatedly that it is important that there be consistent approaches to public access policies to reduce the burden on institutions and grantees. Given the large number of grantees in each research institution, public access policies with standard components and expectations will be important to successful implementation.

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?

Public access with no limitations on use to the final, published article stemming from publicly funded research is preferred but is not always possible. NIH's experience is helpful in understanding the constraints an agency might face in providing access to the final, published article.

NIH mandates the deposit of the final electronic manuscript of a peer-reviewed journal article upon acceptance for publication and that this manuscript be publicly available no later than one year after publication. During the years of congressional consideration of the NIH Policy, some publishers expressed concerns regarding copyright issues. To address these concerns, lawmakers

included language to ensure that the NIH "implement the public access policy in a manner consistent with copyright law."(Division G, Title II, Section 218 of PL 110-161 [Consolidated Appropriations]). As a result, PMC only distributes the copy-edited, published version with the publisher's consent. Authors who deposit manuscripts in PMC retain the appropriate rights to do so thus are consistent with copyright law.

Since the policy went into effect, more and more publishers are collaborating with NIH and are depositing the final, published version. They do this for several reasons: 1) a preference that readers use the publishers' version, 2) it provides a service to their authors, thus there is a competitive advantage, and 3) it may drive more users to their Web site for additional resources. Five hundred and eighty-one journals deposit full journal content to PMC, 178 deposit NIH-funded articles and Springer, Taylor & Francis, Wiley-Blackwell, ACS, APA, the BMJ Publishing Group, and Sage deposit their OA/author-pay articles.

These approaches have responded to publisher concerns by providing a window to protect subscriptions (e.g., access to the final, published version) while at the same time providing public access to federally research results. In addition, it is beneficial if the agency links to the final, published version if the final, published version is not made available to the agency.

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?**

Open access journals, such as those of the Public Library of Science (PLoS), provide immediate, unrestricted reuse and free access to their peer-reviewed journals. Immediate, unrestricted use and free access supports the advancement of science and innovation and ideally would be the preferred approach in implementing public access policies at federal science and technology agencies, particularly given the speed with which science and technology discoveries are made.

This approach, however, could present economic difficulties for some publishers who currently operate under a different marketplace model, the subscription model. As a consequence, most public access policies call for public access to articles stemming from federally funded research following an embargo period of 6 months or less. This is seen with both public and private funders. A comprehensive list detailing these policies is available at the following: http://www.sherpa.ac.uk/juliet/.

There is a range of embargo periods employed by journal publishers. Many make journal articles accessible at 12 months, a growing number at 6 months (over 90 journals published by Nature Publishing Group spanning many

disciplines, for example), and others are more aggressive at 2 or 3 months. Since 2001, the American Society for Cell Biology has provided free access to all of the research articles in *Molecular Biology of the Cell* 2 months after publication. The articles are available on the journal's Web site and in PubMed Central (PMC). Embargo periods for a number of journals are available at the following: http://www.highwire.stanford.edu/lists/freeart.dtl.

Some non-open access publishers have expressed an unfounded concern that immediate access or shorter embargo periods will result in journal cancellations by libraries, as subscription revenue is the primary source of income for most journal publishers. Data has shown that libraries will not cancel subscriptions to journals with shorter embargo periods for several reasons. Researchers, students, and faculty require access to the literature as soon as possible; thus, any embargo constitutes too long of a delay, and journals include needed information and articles well beyond those funded by governments.

The embargo period should not relate to varying levels of access or fair use. Fair use is a key provision in the Copyright Act that is central to the ability of libraries, education institutions, high tech companies and others to achieve their mission and/or to bring new innovative products and services to the market. The federal government is making a policy determination of how, in the words of the RFI, "to leverage Federal investments to increase access to information that promises to stimulate scientific and technological innovation and competitiveness."

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?**

There is a long history of collaboration within the information infrastructure arena to make information accessible in an effective and useful manner. For example, with others, libraries have developed shared systems, standards, and preservation and access strategies to assist users in discovering needed information in all formats. This collaboration is very evident in the development of institutional repositories; many of the needed standards for interoperable, archived, and publicly accessible digital repositories are in place. And as public access policies have been implemented, strategies and standards evolved or developed as required. Another important criteria has been to work with open standards.

With regards to submission format, most agencies and users support a variety of file types, such as MS Word, Excel, and more. Following deposit, some conversion may be required by the repository in order to utilize formats that permit searching, data and text mining, and linking, and today, scientific publishing uses XML. In addition, it will be important for the repositories to

employ a common, standard document type definition, or DTD. Currently, there is broad community support and use of the National Library of Medicine's DTD. Finally, employing Digital Object Identifiers, or DOIs such as PMCID will be important. DOI is a system is for identifying and managing digital content objects and can provide current information, including location on the Internet.

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?**

> *"Science, science policy and the greater public interest all benefit from a culture that is open and transparent as possible. Accordingly, the Federal government should be committted to fostering such an open environment. Office of Science and Technology Policy" Core Principles for the Release of Scientific Research Results," John H. Marburger, Director, OSTP, May 28, 2008.*

Isaac Newton's statement that he "stood on the shoulders of giants" aptly describes how advances in science build on prior knowledge and the sharing of information. Although our investments in cyberinfrastructure have greatly enhanced the exchange of research results and support greater collaboration among scientists around the globe and between scientific disciplines, barriers still remain. Reducing those barriers is essential for advancing scientific discovery, for sustaining economic growth, and spurring innovation in all sectors.

The NIH Public Access Policy provides an excellent example of what can be achieved through the access to software and tools, and the linking of data, databases, journal literature, and researchers. Importantly, it is what users can do with these rich resources that promotes discovery and advances science. Extending this policy to other science and technology agencies would extend useful and effective access to new communities of users with differing interests and perspectives. It would empower these users and support enhanced collaboration across disciplines, nationally and internationally. The Policies must reflect the increasing nature of interdisciplinary, global science. Finally, extending this type of public access policies to other federal agencies will leverage collaborative investments underway, for example the Department of Energy's program to support breakthrough research and the work of the Departments of Housing and Urban Development and Transportation and the Environmental Protection Agency on sustainable communities.

Key to the success of extending public access policies more broadly will be to provide unfettered access to research resources and permit the widest possible use within the law. Utilizing Creative Commons or similar licenses is preferred, and these are widely employed by individuals in all sectors. This will greatly assist in the ability to mine, manipulate, and integrate data and information in publicly accessible digital repositories.

As noted previously, the significant and growing use of PMC indicates the need and value for enhanced access to these federally funded research resources and also demonstrates how useful researchers and others find these tools, software, and databases. Such usage is an important evaluation metric. Removing barriers to scientific communication will translate into new discoveries, including bringing commercial products to the marketplace at a faster pace and for example, as seen through the use of PMC, address pressing national and international health concerns.

For more information, please contact Prue Adler, prue@arl.org.