



## **Public access to federally funded scholarly publications**

SUBMITTED TO THE OFFICE OF SCIENCE AND TECHNOLOGY POLICY, JANUARY 12, 2012

FOR MORE INFORMATION, CONTACT [HEATHER@ARL.ORG](mailto:HEATHER@ARL.ORG)

### **Introduction**

I'm writing today on behalf of the Scholarly Publishing and Academic Resources Coalition (SPARC) in response to the Office of Science and Technology Policy's Request for Information (RFI) dated November 3, 2011, seeking input on the issue of Public Access to Scholarly Publications resulting from federally funded Research.

SPARC is an international alliance of more than 800 academic and research libraries that promotes expanded sharing of scholarship in the networked digital environment. SPARC believes that wider and faster outputs of the research process increases the impact of research, accelerates the pace of scientific discovery and promotes innovation and increased return on our collective investment in research.

SPARC was formed to act on the library community's desire to ensure that the promise of the Internet to dramatically improve scholarly communication, particularly in the journals marketplace, was realized. It has been an innovative leader in the rapidly expanding international movement to make scholarly communication more responsive to the needs of researchers, students, the academic enterprise, funders, and the public. Its pragmatic agenda focuses on collaborating with other stakeholders to stimulate the emergence of new scholarly communication norms, practices and policies that leverage the networked digital environment to support research and expand the dissemination of research findings.

We appreciate the leadership the Administration has shown in promoting openness in the federal government, and share the belief that policies are needed to implement the principles of transparency, participation, and collaboration, as set forth in the Open Government Directive of 2009. We thank the Office of Science and Technology Policy for convening this substantive discussion on the importance of ensuring broad public access to the results of federally funded research. SPARC shares the Administration's view that enhancing access to this information will promote advances in science and technology, encourage innovative use and application of government-supported research, and fuel commercial development and economic growth.

This RFI was issued in accordance with the America COMPETES Reauthorization Act of 2010, legislation that is aimed at improving the competitiveness of the United States through investment and innovation in research and development. We firmly agree that

an effective government-wide policy to ensure public access to the results of the ~\$60 billion<sup>1</sup> in scientific research conducted annually using public funds is an essential component of achieving the goals of the America COMPETES Act.

We support the establishment of a public-access policy covering all federal agencies that conduct scientific research. To maximize the returns on our nation's investment, such a policy should ensure that all members of the public are able to immediately access and fully reuse digital articles reporting on the results.

SPARC views the successful National Institutes of Health (NIH) Public Access Policy<sup>2</sup> – which requires NIH-funded researchers to make the final authors manuscript freely accessible to the public through the agency's permanent, digital archive (PubMed Central) no later than one year after publication in a peer reviewed journal – as an appropriate baseline policy, with one important modification. To maximize the impact of federally funded research, any government-wide public-access policy must ensure that digital articles reporting on the results can be fully used by all downstream users (scientists, teachers, entrepreneurs, etc.). We support ensuring this through the use of a license that works within the current copyright system that at most requires attribution to the author, such as the Creative Commons Attribution (CC BY) license.

Our responses to the specific questions in the RFI follow below.

**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 publicly 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?**

#### *Encouraging Commercialization*

As noted earlier, this RFI was issued in accordance with the America COMPETES Reauthorization Act of 2010, legislation designed to improve the competitiveness of the United States through investment and innovation in research and development. To compete effectively in a global economy, the U.S. must ensure that it is positioned to translate ideas generated from publicly funded scientific research into new services and products as rapidly as possible.

Articles reporting on the results of publicly funded research are a critical component of our nation's research output. To create the optimal environment to encourage such commercialization, the complete collection of full-text<sup>3</sup> articles reporting on federally funded research should be made immediately, freely available to the public. Members of

---

1 <http://scienceprogress.org/2011/07/u-s-scientific-research-and-development-202/>

2 The NIH reports that more than 500,000 unique individual users access the PubMed Central Database daily to view the more than 2 million full text articles it contains.

3 It is important to note that the full text of peer-reviewed articles must be made accessible – not merely abstracts, summaries, or un-peer reviewed grant reports – for the full value of these articles to be leveraged by the public.

the public must also be ensured the rights to fully use these articles (i.e., to text and data mine, compute on, etc.) without commercial restriction. Enabling immediate access and full reuse – a concept known as Open Access<sup>4</sup> – to these articles will accelerate the ability of individuals and companies to construct new services and products, and ensure that the value of the public's investment in this research is fully realized. It will create a business climate where all stakeholders – in both the public and private sector – can incorporate ideas generated from this research into their business development cycles more quickly, speeding the launch of new products, and accelerating the creation of new markets.

There is a substantial body of research that provides evidence of the tangible economic benefits of increased access to this information, specifically in terms of product innovation and increased revenues through sales.<sup>5</sup> It is important to note that the full text of peer-reviewed articles must be made accessible – not merely abstracts, summaries, or un-peer-reviewed grant reports – for the full value of these articles to be leveraged by the public.

Providing fast, barrier-free access to articles allows more users to stay abreast of cutting-edge ideas, and generate new uses and applications for research. This is particularly crucial for startups, and small- and medium-sized businesses which frequently find the cost of accessing these articles (whether through subscriptions or pay-per-view models) prohibitively expensive. In one study examining the ability of such businesses to access scientific articles, over a third of the companies reported they frequently had trouble accessing articles, while an additional 41% reported that they sometimes experienced difficulty.<sup>6</sup>

This lack of access directly impacts these companies' ability to be competitive. One study estimated that, without a mechanism to ensure open access to articles reporting on the results of research, it can take a small business approximately two years longer to be able to incorporate ideas generated from it into their work streams<sup>7</sup> – an unacceptable lag time if the goal is to speed our nations' ability to compete effectively in a global economy.

This research also highlights a key advantage to making articles resulting from federally funded scientific research openly accessible to businesses for commercial use. While academic users of research articles often take a discipline-specific approach, businesses rarely do. Rather, business use of research is often characterized by cross-disciplinary and multidisciplinary work. Subscription costs often present insurmountable barriers for businesses – particularly startups and small businesses – to effectively exploit this research.

The experience of funding agencies (such as the National Institutes of Health) that have implemented a mandatory public-access policy has demonstrated that the public has a deep and abiding interest in articles resulting from publicly funded research, and in the

---

4 <http://www.soros.org/openaccess/read>

5 See, for example, Narin et al (1997), Mansfield (1995), Toole (1999), Tijssen (2002), McMillian and Hamilton (2002).

6 Houghton, Swan and Brown (2011)

7 Houghton, Swan and Brown (2011)

creation of tools and analytics that can provide new insights into this research.<sup>8</sup> Their experience has also demonstrated that such policies create opportunities that are complementary to, rather than competitive with, the current journal publishing marketplace. During the nearly four years that the NIH Public Access policy has been in place, the Science, Technology and Medical (STM) journal industry has reported increases in growth and revenues for its subscription access journals.<sup>9</sup> We know of no data that has been reported by any publisher indicating that they have been negatively financially impacted by the policy.

Over the same time period, a robust new market of open-access journals – journals that make their content freely available to all users, with no restrictions for reuse other than appropriate attribution to the author – has also flourished. More than 7,300 open-access journals are currently being published in a broad spectrum of disciplines.<sup>10</sup> Innovative companies, such as the U.S.-based Public Library of Science (PLOS), have lead the way in demonstrating the financial viability – and desirability – of this new publishing model. The successful establishment of the innovative and profitable PLOS One journal, now the single largest scientific journal published in the world,<sup>11</sup> quickly spurred the introduction of similar products by both for profit and non-profit publishers.

In fact, the majority of large commercial publishing companies have moved rapidly to invest in the development of their own open-access journal publishing programs, with major players such as Springer, Wiley, Sage, Nature, and Taylor and Francis all rolling out new open-access options over the past several years. The growth of this new market segment has been so dramatic that a new trade association, the Open Access Scholarly Publishers Association (OASPA) has now been established to help promote its further development.

The emergence of this growing collection of openly accessible publications has already begun to spur development of new tools and services. Opportunities abound for innovation in search, data and text mining, current awareness, integration of data, impact measurement, translations, indexing, recommendations and summarizing. Mendeley, a company launched in 2009, offers integrated academic search and peer recommendations entirely built on a collection of open-access papers. This rapidly growing company already boasts a user base number over one million individuals. Companies that build applications that mine publication databases and produce meta-analyses and visualization of the pattern of scientific results from multiple studies (such as Pub Gene and BioCreative) are also benefiting from availability of more open publications.

Similarly, companies such as iParadigm, which builds tools used to detect plagiarism, routinely mine open-access databases – including PubMed Central – as one of their primary resources to power their iThenticate tool.

---

8 <http://olpa.od.nih.gov/hearings/111/session2/Testimonies/PublicAccess.pdf>

9 [http://www.stm-assoc.org/wp-content/uploads/outsell\\_december\\_2011\\_large.jpg](http://www.stm-assoc.org/wp-content/uploads/outsell_december_2011_large.jpg), STM Publishing News, <http://www.stm-publishing.com/?p=722>

10 <http://www.doaj.org>

11 [http://en.wikipedia.org/wiki/PLOS\\_ONE](http://en.wikipedia.org/wiki/PLOS_ONE)

The establishment of a government-wide public-access policy for articles generated by federal science agencies other than the NIH will help further fuel development of this burgeoning new market area. Far from threatening the publishing industry, publishers are well positioned to lead the development of many of these new services.

While creating new markets and business models in the publishing industry is one important outcome of a public-access policy, it is important to remember that it is part of a larger goal: to encourage the use of publicly funded research to spur increased commercialization in other business sectors. Creating a government-wide policy that results in an openly accessible database (or set of databases) of publicly funded articles will provide opportunities for companies of all kinds to build on this information, in a manner similar to services that have been built on other public data, like that generated by the National Weather Service. An effective policy will encourage additional private investment in IT to capitalize on a government resource, a proven strong point in the U.S. economy.

For many industries (such as the biotechnology and pharmaceutical industries) the ability to interact with leading-edge research results is part of the lifeblood of the company. They (and their investors) count on these resources to be able to deploy a research and development strategy that keeps them on the cutting edge of new ideas and knowledge, so that they can translate these ideas quickly into marketable products and services. A growing body of evidence demonstrates that the pace and volume of commercialization is markedly increased with access to a database of openly accessible – and fully reusable – content, than with access to one that applies access or use restrictions.

For example, a recent study compared the commercial outputs deriving from the openly accessible Human Genome Project database versus the restricted access Celera genome database. The study showed that the number of products (specifically, gene-based diagnostic tests) developed from the open Human Genome database was approximately 30% higher than from the restricted access Celera database. Perhaps most notably, the study also found that this commercial advantage persisted over time, even after access and usage restrictions were lifted from the Celera database.<sup>12</sup>

This finding is supported in the 2011 report "Benefits to the Private Sector of Open Access to Higher Education and Scholarly research," an extensive examination of the potential impact of open access on businesses by the Joint Information Systems Committee (JISC) in the U.K.<sup>13</sup> The authors examined how a government policy supporting open access to research articles reporting on publicly funded science might effect companies of various sizes in a wide range of disciplines – from the pharmaceutical powerhouse GlaxoSmithKline to Rolls Royce, to smaller microelectronic companies and architecture firms. Positive impacts reported centered first and foremost on the (not insignificant) time and cost savings to the companies in time spent searching for relevant articles, and constructing workarounds to pay walls or copyright restrictions. However, many of the business also reported specific examples that illustrate potential wider benefits of open access. These included:

---

12 Intellectual Property Rights and Innovation: Evidence from the Human Genome Project, Heidi Williams  
13 [http://open-access.org.uk/wp-content/uploads/2011/10/OAIG\\_Benefits\\_OA\\_PrivateSector.pdf](http://open-access.org.uk/wp-content/uploads/2011/10/OAIG_Benefits_OA_PrivateSector.pdf)

- the location of previously unknown experts or information in unrelated disciplines which lead to breakthroughs in new product or service developments,
- the development of new research and business areas previously not considered viable, and
- the establishment of new partnerships with universities housing open-access repositories, among others.

#### *Improving Scientific Productivity*

Besides providing an environment in which commercialization can be optimized, ensuring full open access to articles reporting on the results of publicly funded research can also play an important role in improving scientific productivity. To do this, we have to move away from thinking about access to articles as simply something considered at the end of the research process. Rather, ensuring the ultimate accessibility and utility of articles is a critical part of the design of the research process, and should be taken into account at the start. This reflects an important reality in scholarly research: that the work is not complete until the results have been fully communicated and are openly available for others to build upon. Public-access policies can play a crucial role in ensuring that our nation's scientific research infrastructure is designed to optimize the accessibility and utility of these articles from outset, amplifying all of the desired outcomes from publicly funded research.

The research community has long recognized the opportunity that providing immediate, barrier-free, online access presents to researchers to work faster, by enabling them to get to research articles and incorporate new findings into their research more rapidly. In biomedical disciplines, for example, the need to rapidly collect, evaluate and understand the work of colleagues is readily apparent. In 2009, the very immediate public health threat presented by the influenza virus (H1N1) highlighted the need for the rapid exchange of scientific results and ideas in this arena. Finding traditional journal channels inadequate to the task, the community of expert influenza researchers self-organized, and created a new, open-access online resource for the immediate, open communication and discussion of data and ideas in the field, PLoS Currents: Influenza.<sup>14</sup> This new open-access resource quickly gained the support of the community, and rapidly evolved into a crucial piece of the influenza research landscape.

Ensuring open access to scientific articles can also help scientists to incorporate more information into their work more efficiently. As the amount of digital information continues to expand at a breakneck speed, it is crucial for researchers to be able use new tools to access, read, and understand a larger amount of information in a shorter amount of time. One of the most exciting aspects of enabling an open-access environment is that it provides a research infrastructure in which computers are fully enabled as new category of reader. Machines can help researchers work much more efficiently, powering through huge numbers of digital articles that a human reader alone has no possibility of digesting.

With the continued increase in papers generated from scientific research, enabling computers in this way is an essential element in improving scientific productivity.

---

<sup>14</sup> <http://knol.google.com/k/plos-currents-influenza>

Consider the case of biomedicine – there are currently more than 19 million citations and abstracts covered by the National Library of Medicine's search engine, PubMed. These include ~830,000 articles published in 2009, up from 814,000 in 2008 and 772,000 in 2007.<sup>15</sup> The growth rate gives no indications of slowing, particularly as emerging economies like India, China and Brazil continue to accelerate their research outputs.

An open-access environment also allows researchers to employ new semantic and computational tools that can help scientists to contextualize ideas contained in papers, identifying new relationships and significantly expanding the breadth of research threads. Studies have demonstrated that open access to research may contribute to the exploration and creation of novel research lines, increasing the diversity of experimentation that follows from a single idea.<sup>16</sup> This is key in accelerating discovery, as progress in science is not a linear function.

In Alzheimer's research, for instance, scientists can be overwhelmed by the sheer volume of papers on the disease. It is difficult to keep up with the volume, and to accurately assess whether a new hypothesis is consistent with the existing literature. This led to the creation of tools such as the Semantic Web Applications in Neuromedicine (SWAN), a new tool designed to help researchers locate the papers most relevant to their interests, and uncover connections that might not otherwise be obvious – as well as to test and generate new hypotheses.

SWAN is a curated, online repository of hypotheses on Alzheimer's research derived from published articles, and provides a visual, color-coded display of the relationships between the hypotheses, highlighting overlaps, agreements, and conflict. Researchers have noted that it helps them to advanced their research – not only helping to focus it in a certain direction, but also to broaden it to include others.<sup>17</sup> Ensuring that these tools can fully function, using the full text of papers reporting on the results of publicly funded research, would greatly enhance the value of the research to the public, and significantly enhance the productivity of the research community.

A government-wide policy that facilitates the creation of an open-access environment will also allow – and encourage – more people to participate in the scientific research process at many levels. Researchers in a variety of disciplines are already using open-access environments (for both data and publications) to help them expand their pool of collaborators in specific research areas, as well as to help create new pathways to solutions.

In another example from the field of Alzheimer's research, experts (led by Neil Buckholtz, chief of the Dementias of Aging Branch of the Division of Neuroscience at the U.S. National Institute on Aging, and Dr. William Potter, a neuroscientist at Eli Lilly) established the Alzheimer's Disease Neuroimaging Initiative (ADNI), a novel, public-private collaboration that posts all of its data on Alzheimer's on an open public Web site. ADNI has made thousands of brain scan images and clinical and neuropsychological data

---

15 <http://www.nature.com/news/2010/100127/full/463416a.html>

16 Murray et al, "Of Mice and Academics" <http://www.nber.org/papers/w14819>

17 Mureson, <http://www.nature.com/news/2010/100127/full/463416a.html>

available to researchers around the world, and has generated a wealth of new research papers, as well as more than 100 new studies testing drugs that may slow or stop the disease.<sup>18</sup> The ADNI model is already being replicated other areas, most notably in Parkinson's disease research.<sup>19</sup>

Along with increasing the sheer number of participants, an open-access research environment also increases the diversity of participants in the research process. It helps to promote access and reuse of information by researchers in loosely related (or even unrelated) fields that might not otherwise have access to the full corpus of research articles. This increases the value of our scientific research investment, by increasing the efficacy of scientific discovery.

For example, one study by Lakhani et al. measured the impact that reaching out to communities of researchers beyond a primary research domain had on research problem-solving. They found that the inclusion of researchers from outside fields resulted in their ability solve one third more problems than experienced R&D firms were able to solve on their own.<sup>20</sup>

This diversity of participants also translates into increased opportunities for collaboration with entire communities who have a stake in the outcome of research – such as the patient's advocacy community. Providing immediate, free access to articles that report on the latest results of publicly funded research can facilitate a fundamental change in the way stakeholders – parents, providers, and advocates – are included in the discovery process.

In the Autism community, Sophia Colamarino (Stanford University Medical School, and former Vice President for Research at Autism Speaks) has spoken eloquently on the patient advocacy and researcher funding perspectives, pointing out why public access is specifically important for patient groups. She notes that, because there is no routine treatment for Autism, families are routinely responsible for learning about therapies and treatments that may be appropriate for them. She points out that these families consistently seek reliable information, and that they are sophisticated in their ability to read and interpret scientific literature.

She further notes that her experience has shown that while families are inundated with information from a variety of sources, what is most easily available may not always be credible. Because of the barriers that subscription and pay-per-view pay walls present, families have easy access to all BUT the most scientifically valid information.<sup>21</sup>

Providing immediate, barrier-free access to articles that report on the results of publicly funded research access empowers family members and caregivers to be better, more informed advocates, and gives them a positive outlet by allowing them to participate in

---

18 <http://www.nytimes.com/2010/08/13/health/research/13alzheimer.html>

19 [http://www.michaeljfox.org/living\\_PPMI.cfm](http://www.michaeljfox.org/living_PPMI.cfm)

20 Lakhani, "Marginality and Scientific Problem Solving Effectiveness through Broadcast Search," *Organization Science* 21 (September - October 2010): 1016-1033.

21 <http://www.berlin9.org/bm~doc/berlin9-colamarino.pdf>

progress first hand. Barriers to accessing published research literature cause families to struggle to find the most rigorous data necessary to make informed decisions.

### *Costs and Benefits*

#### Costs

Understanding the potential costs and benefits of ensuring that articles reporting on publicly funded research are made accessible is of critical importance, and is an area where many helpful sources of data are available to draw on. To explore the potential costs, it is useful to examine the data provided by the National Institutes of Health (NIH), whose successful public-access policy already ensures full accessibility to articles reporting on the results of the ~\$30 billion of basic and applied research that it funds annually.

This policy, which covers approximately one half of the total U.S. annual investment in scientific research, has proven to be extremely cost-effective. NIH reports that it costs \$3.5- \$4.6 million annually (on a total \$30-billion budget) to administer its public-access policy. This represents an investment of only about 1/100th of one percent of the NIH's overall \$30 billion operating budget to ensure that the 90,000-95,000 articles generated annually to report on NIH-funded research are readily accessible to all potential users.<sup>22</sup>

The NIH also reports a deep demand for these articles, with more than 500,000 unique users from all sectors of the public accessing the PubMed Central database each day to view and retrieve articles.<sup>23</sup> An effective, government-wide public-access policy can likewise be implemented in a cost-effective manner, by leveraging this existing infrastructure to minimize unneeded duplication of efforts, and utilizing the investments already made by the NIH.

#### Benefits

In 2005, the Organization for Economic Development explicitly noted that “Governments would boost innovation and get a better return on their investment in publicly funded research by making research findings more widely available... and by doing so, they would maximize the social returns on public investments.”<sup>24</sup>

It is no surprise that the concept of ensuring open access to the results of publicly funded research is now a cornerstone of innovation and competitiveness policies around the world.<sup>25</sup> As Neelie Kroes, Commissioner for the Digital Agenda of the European Commission recently noted, “Open access to scientific information is not a luxury; it is a must if we want to compete globally.”<sup>26</sup>

Significant economic research has been done, in the U.S. as well as internationally, on cost-benefit analyses of various policy approaches to ensuring greater access to articles

---

22 <http://www.hhs.gov/asl/testify/2010/07/t20100729c.html>

23 <http://olpa.od.nih.gov/hearings/111/session2/Testimonies/PublicAccess.pdf>

24 <http://www.oecd.org/dataoecd/42/12/35393145.pdf>

25 <http://www.roarmap.eprints.org>

26 Neelie Kroes video <http://www.youtube.com/watch?v=taux110Vgek>

reporting on the results of publicly funded research. Detailed economic analyses have been conducted on proposed national policies in Australia, the U.K., the Netherlands and elsewhere, providing sound methodologies for policy makers to use in considering the potential impact of such policies. These studies have consistently demonstrated that the adoption of policies to encourage the open sharing of research results – including scientific articles – has a significant economic upside for national economies.<sup>27</sup>

Perhaps most germane for the purposes of this RFI is the 2010 study conducted by Houghton et al., examining the potential impacts of opening up access to articles reporting on the results of all U.S. federally funded scientific research, under a policy similar to that of the current NIH Public Access Policy. Houghton and his colleagues examined both the costs and potential returns to the public investment in R&D, and provide a working model to be used for further testing and refining estimates as additional data becomes available.<sup>28</sup>

The initial Houghton et al. modeling suggests that providing open access to all articles reporting on U.S. scientific research under a model similar to the current NIH policy would (very conservatively) result in at least a five-fold increase in ROI, with the benefits of the policy estimated to be approximately 8 times larger than the costs. They further estimate that the net present value gains of expanding an NIH-style policy to all other U.S. science agencies over time would be on the order of \$1.5 billion (net the costs of running the archive). Of that number, approximately 60% is estimated to accrue directly to the U.S. economy.<sup>29</sup>

#### *Accountability and Efficiency*

A government-wide public-access policy will have the added benefit of supporting informed, transparent, science-based federal budget and policy decision making by increasing federal agency accountability and providing agencies with an improved accounting on the outcomes of their funded research. It will also give Congressional budget drafters, appropriators, and authorizers better information to accurately assess the value of existing expenditures, and to target funding on the most promising research areas.

#### *What Type of Access is Needed?*

To maximize U.S. economic growth and improve the productivity of the American scientific enterprise, the complete collection of full-text articles reporting on federally funded scientific research should be made immediately, freely accessible to the public. The public must also be ensured the rights to fully use these articles (i.e., text and data mine, compute on, etc.) without commercial restriction.

---

27 <http://www.cfses.com/projects/Easi-OA.htm>

28 Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs, Houghton et al. (2010)

29 Op. cit.

**2. What specific steps can be taken to protect the intellectual property interests of publishers, scientists, federal agencies, and other stakeholders involved with the publication and dissemination of peer-reviewed scholarly publications resulting from federally funded scientific research?**

To best support the goals of accelerating scientific discovery, innovation and the creation of new markets, any public-access policy should ensure not only full accessibility of scientific articles, but also full utility of the articles in the digital environment. Mechanisms to enable full use (i.e. distribution, reuse, text mining, data mining, computation, etc.) should be part of any government-wide public-access policy. Digital articles reporting on publicly funded research should contain a clear description of the rights and permissions granted to users – both human and machine readers. A policy that results in a "read-only" database severely limits the utility – and the value – of the information contained in these articles.

Public-access policies can be successfully implemented while respecting and working within the current copyright framework. A policy requiring full open access to articles reporting on the results of federally funded research under a mechanism such as the Creative Commons Attribution (CC BY)<sup>30</sup> license is consistent with protecting the copyrights of both authors and publishers. Such a license does not replace copyright protection, but rather works within the copyright system, giving authors and publishers the ability to mark their content with the rights available, while ensuring that the rights holder receives credit in whatever manner they prefer. This type of license also serves the needs of the author by enhancing their ability to share their work more widely, increasing the opportunities for their work to be used and cited.

While the NIH Public Access Policy provides an excellent benchmark for most aspects of government-wide policy, one where it can be substantially improved upon is in the area of clarifying rights retention. The Department of Labor's Trade Adjustment Assistance Community College and Career Training (TAACCCT)<sup>31</sup> grant program provides a more appropriate exemplar. The TAACCCT program requires that grant recipients license content created from grant funds under a Creative Commons Attribution (CC-BY) license. This framework ensures broad access and reuse for anyone wishing to utilize this federally funded research output, while also ensuring that proper credit is given to the author.<sup>32</sup>

The public also needs full use of these articles sooner than the current term of copyright allows. Ideally, articles reporting on the results of publicly funded research should be made accessible to the public immediately upon appearance in a journal. However, an initial interim, phased approach might prove a practical way forward. This type of approach might be constructed to include:

---

30 <http://creativecommons.org/licenses/by/3.0/>

31 <http://www.doleta.gov/grants/pdf/SGA-DFA-PY-10-03.pdf>

32 <http://epsiplatform.eu/content/topic-report-no-23-creative-commons-and-public-sector-information-flexible-tools-support-psi>

- First, providing an appropriate period of embargoed access (no longer than 12 months) where current rights appropriate under copyright apply;
- Second, after the expiration of the embargo period, full reuse [consistency] rights under an appropriate license such as CC-BY apply.

It should be the explicit goal of any government-wide public-access policy to make the results of federally funded research as useful as possible. Broad reuse allows both researchers and businesses to unlock additional value from our public research investment – now, and for decades to come. Restrictions that limit how users can work with these digital articles will result in only a fraction of their value being delivered, and unnecessarily reduce the subsequent return to the taxpayer.

### **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?**

The federal government is the appropriate entity to provide permanent stewardship of these articles, and is in a unique position to ensure that publicly funded articles are made permanently accessible, and useable. To ensure this, any public-access policy that is developed must give the federal government adequate rights to archive and distribute articles reporting on publicly funded research.

There are many compelling reasons for the federal government to maintain custody of articles arising from federal research. The library community fully understands that while multiple copies are necessary to support archiving and preservation, there also should be one master copy that is maintained by a recognized leadership organization. Currently, the National Library of Medicine (NLM) appropriately fulfills that crucial role for articles generated by NIH-funded research by housing articles in their PubMed Central (PMC) digital repository.

NLM has indicated<sup>33</sup> that they are willing to expand their role and accept articles from any other federal science agency, providing an immediate, cost-effective potential solution. Alternatively, NLM has also indicated that the software supporting PubMed Central is freely available in the public domain, and was explicitly designed in a modular form to be easily shared with other entities that might wish to use it. This option provides another cost-effective mechanism that ensures the interoperability of multiple federal agency archives. This solution has also been demonstrated to be effective, and is currently in use by the Canadian Institutes of Health Research<sup>34</sup> as well as the Wellcome Trust<sup>35</sup> in the U.K.

Having the federal government retain custody of a master copy of these articles also minimizes the possibility that any entity can create an exclusive arrangement that would inhibit – if not eliminate – the ability for a wide variety of stakeholders and businesses to

---

33 <http://olpa.od.nih.gov/hearings/111/session2/Testimonies/PublicAccess.pdf>

34 <http://www.cihir-irsc.gc.ca/e/34846.html>

35 <http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/UKPMC-Project/index.htm>

maximize the utility of these articles, and ensure that new services and products can be readily built from them. It also ensures that overall policies and practices, including access conditions another standards, can be identified and applied effectively across multiple archives if needed. It is crucial that any federal archive be an access point for the public, and not simply serve as a "dark" archive. The library community's long experience with archiving publications has clearly shown that regular access and use of digital materials is a crucial element in effective long-term preservation.

This type of approach does not preclude other, non-governmental entities from also participating as partners in a decentralized approach. An effective federal public-access policy could also involve multiple repositories maintained by third parties, as long as those repositories support access and use conditions that allow all interested parties to build on the content contained in them. Repositories that meet conditions for public accessibility, unrestricted use rights, interoperability and long-term preservation of articles can play an important role, encouraging innovative public-private partnerships.

Some have posited that it is a duplicative cost and effort for the federal government to maintain an accessible archive of these articles, citing efforts by publishers in the private sector, such as CLOCKSS, Portico and others. While we support the important efforts of these organizations and projects, they do not provide sufficient coverage to ensure that the entire corpus of articles generated through public funds are made available. For example, a recent examination of archival coverage of journal articles by two major research university libraries – Cornell University and Columbia University – indicated that, after years of active participation in such projects, only about 15% of their journal holdings are currently archived by LOCKSS and Portico combined.<sup>36</sup>

Federal government stewardship is a necessity to ensure that our investment in scientific research is protected and leveraged now and into the future. As discussed in detail in Comment 1, federal stewardship is also cost-effective. With over a decade of experience in running the PubMed Central archive at the National Library of Medicine, the NIH has demonstrated that a small investment can ensure that the public has full access to the results of the scientific research that it collectively funds.<sup>37</sup>

#### **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?**

Public-private partnerships can play an important role in leveraging the unique capabilities of a broad range of potential service providers, and create opportunities for the development of new products and services to built on publicly funded information. A key aim of the America COMPETES Act (whose goals this RFI has been issued to facilitate achieving) is to improve the competitiveness of the United States through investment in research and development. As such, it is critical that any public-private

---

36 <http://2cul.org/node/22>

37 <http://olpa.od.nih.gov/hearings/111/session2/Testimonies/PublicAccess.pdf>

partnerships be constructed to ensure that *all* potential service providers have an equal opportunity to participate. Under no condition should any one site, organization or company be the single point of access for publicly funded articles.

This is particularly important as it relates to small businesses that may experience difficulty with entering markets given access conditions or restrictive copyright/reuse provisions. Constructing a partnership that unfairly advantages a limited number of participants will result in a less competitive environment, rather than facilitating the kind of environment that encourages robust participation as envisioned in the COMPETES Act.

The publishing community should be encouraged to participate in such partnerships, but it is only one stakeholder group whose interests must be considered. The federal government should also carefully consider the synergies in mission that exist with other potential partners, particularly libraries, archives and museums. These entities have missions explicitly focused on long-term preservation and access to information, and also have a wealth of experience and existing infrastructure that can be leveraged. Developing a public-access policy that includes roles for these kinds of organizations would greatly increase prospects for the viability and long-term sustainability of such partnerships.

It is also notable that none of the 50+ research funders that currently have active public-access policies are using proprietary or publisher-based sites as their final archives.<sup>38</sup> However, there are good examples of funders partnering with academic and research institutions in this role. The Digital Repository Infrastructure Vision for European Research (DRIVER) project is one such example.<sup>39</sup> DRIVER has constructed a pan-European infrastructure for digital repositories, providing a linked network of accessible archives to house the results of publicly funded research (both digital data and articles) for use by researchers, administrators, and the general public.

Public-private partnerships should be encouraged, provided repositories meet conditions for public accessibility, use rights, interoperability and long-term preservation of publicly funded articles, ensuring that all stakeholders and businesses have opportunities to participate effectively.

**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-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to federal science funding?**

Metadata plays a crucial role in enabling the interoperability, search, discovery and analysis of articles reporting on federally funded research. Rather than thinking about metadata as simply a description of items contained in repositories, the federal

---

38 <http://roarmap.eprints.org>

39 <http://www.driver-repository.eu/>

government should view it as a means to enable specific actions that can be taken on digital articles, as well. To be as useful as possible, metadata associated with federally funded articles must be both machine-readable and machine-interoperable, and should facilitate the robust use, reuse and analysis of digital articles.

While it is possible to identify, as requested, a "minimum core" set of metadata (Dublin Core/OAI-PMH would be the current de facto standard, coupled with the use of the NLM DTD for article markup), this will only facilitate a minimum amount of discovery and download. Broader metadata specifications are needed to make full use of the information contained in federally funded articles and to active the aims of the federal government of improving scientific productivity and accelerating commercialization.

To maximize the value of this information, additional metadata is needed to also facilitate archiving and preservation, and to encourage the development of new services (such as text mining, visualizations, etc.). This metadata could include:

- Persistent identifiers for authors, publications and links to data
- Controlled identifiers, such as ORCID, 1/2
- Usage tracking/analytics, such as COUNTER, SUSHI
- Metadata supporting context for published resources (i.e., controlled vocabularies)
- Attribution for funding organizations
- Grant IDs
- Descriptions of relationships between entities (such as RDF and OWL enable)

It is important to ensure that any metadata standard or framework not only meet current needs, but also be flexible and extensible enough to support potential future uses. This is particularly critical to ensure that connections between articles and digital data can be supported.

Close consultation with established entities that are working on standards and best practices in this area, such as NISO and the Library of Congress, will also be helpful and should be actively pursued.

## **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?**

Federal agencies that fund science can maximize the benefit of public-access policies to U.S. taxpayers by making the complete collection of full-text articles reporting on federally funded scientific research immediately, freely accessible to the public. Members of the public also must be guaranteed rights to fully use these articles without commercial restriction. The federal government should provide long-term stewardship over the repositories that house these articles, in partnership with organizations such as libraries and archives. Access conditions and reuse rights that at most require author attribution (Creative Commons Attribution CC-BY license or similar) must also be clearly articulated, to enhance scientific productivity and encourage the full range of potential

stakeholders to build secondary services and generate new products and markets from this content.

For any public-access policy to be successful, there must be consistency of requirements across all federal agencies. Most federal grants are managed on behalf of researchers by their home institutions, which, in turn, host researchers who hold grants from multiple funding agencies. Creating disparate access policies – or even compliance requirements – for different federal science agencies would introduce needless confusion and expense into the system, and greatly increase the compliance burden on the grantee and their home institution. Uniform requirements and procedures regarding deposit of peer-reviewed articles should be established across all federal agencies covered by a public-access policy to reduce the cost and complexity of compliance.

Effective implementation strategies that minimize the burden on the researcher can also play an important role in maximizing the returns to the taxpayer, by raising compliance rates and ensuring that the complete corpus of articles reporting on federally funded research is widely available in a timely manner. For example, any government-wide public-access policy should be constructed to take advantage of existing protocols to facilitate automatic deposit of articles to multiple repositories, such as the SWORD protocol.

To optimize the benefits to the taxpayer, any federal public-access policy should also be constructed in a way that encourages the development of additional tools and services to facilitate both the work of the researcher, and the federal agency. Encouraging the integration of articles with agency (and home institution) grant management systems is an important potential way to improve agency accountability, as well as to provide increased information to the public on the results of the research that their tax dollars support.

An effective public-access policy centered on creating accessible databases of research articles can also create opportunities to build productivity management tools - like enhanced bibliographies or Principle Investigator (PI) Profiles – that are of wide use to researchers, institutions, and federal agencies. Similarly, policies based on the creation of such an openly accessible database might provide new opportunities for higher education institutions to help raise the visibility of research arising from their campuses. They also might provide an opportunity for such a database to be used as a teaching tool – perhaps providing an environment to teach researchers and scholars new techniques for effective literature search and analysis.

**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?**

The principle of the public's right to access to articles that report on the results of publicly funded research applies to all outputs of research. Data and educational materials (book chapters, texts, conference proceedings, etc.) that result from publicly funded research should also be made readily accessible to the public.

However, we recognize that different conditions and expectations apply to different types of outputs. For example, authors are not paid for journal articles, but may in fact be compensated for the creation of book chapters. Data sets may contain confidential or personal information that may not be appropriate for unrestricted access or reuse. Access policies that reflect these differences while holding true to the basic principle of public access may need to be constructed.

**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 weigh public and private benefits and account for external 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?**

To optimize their scientific and commercial utility, articles reporting on the results of federally funded research should be made immediately available to the public in freely accessible digital repositories. The federal government should also consider providing support to cover reasonable publication fees for those authors who opt to publish their articles full open-access journals (those that are immediately freely accessible, and enable full reuse rights such as those supported by the Creative Commons Attribution (CC-BY) license). However, to accommodate those journal publishers that continue to rely on subscription income, an author-determined embargo period that is as short as possible – preferably 6 months, but certainly no longer than 12 months – could be considered.

Some publishers have argued that public-access policies – including the NIH Public Access Policy, which include a lengthy 12-month embargo period – will discourage their primary customers (academic libraries which also constitute SPARC's primary member base<sup>40</sup>) from continuing to subscribe to journals and cause them financial harm. Any data documenting such a negative impact should be carefully considered; however, we know of no studies that directly examine this hypothesis nor any documented examples of journals whose financial viability has been significantly damaged by public-access policies.

SPARC's academic and research library members affirm that, while significant journal cancellations are taking place, they are consistently driven by a combination of annual journal subscription price increases, coupled with library budget reductions. These data are well-documented, and published annually in multiple publicly available outlets.<sup>41</sup>

Given the rapid growth in the number of open-access journals, and the increasing adoption of the open-access model by publishers across the journal marketplace, we note that the use of embargoes only benefits one subset of publishers that use a very specific, subscription-dependent revenue model. Open-access publishers, whose business models replace subscription fees with article processing fees, institutional subsidies, advertising,

---

40 <http://www.arl.org/sparc/member/>

41 <http://www.outsellinc.com/insights/11448>

and other revenue streams, have very different revenue models, and receive no clear benefit from embargoes of any length.

The discussion of the inclusion of embargoes in public-access policies often centers exclusively on their potential to protect publisher revenues. However, since one of the goals of an effective federal public-access policy is to balance the needs of all stakeholders, it is also important to consider the impact of embargoes on other stakeholders. Embargoes of any length come with a cost in terms of decreased public access and a negative impact on the degree to which an article's availability fosters further research and development.<sup>42</sup>

In examining the length of embargo periods currently in use, a maximum embargo period of six months has emerged as the norm among biomedical research funders, with the NIH an outlier allowing 12 months. In other disciplines, embargoes of maximum 12 months are most prevalent in research funder policies around the globe.<sup>43</sup> This is also consistent with the current voluntary practices of many publishers. Highwire Press, one of the premier online hosting services for scholarly journals, currently lists hundreds of journals in a variety of disciplines that make their articles freely accessible after a 12-month (or shorter) embargo period.<sup>44</sup>

If different embargos are to be considered for different disciplines, the full range of factors that affect each discipline must be taken into account. These factors include at minimum:

- Growth of journals and papers in discipline (competitive analysis)
- Price – and pricing history – of journal and competitive titles
- Impact of required bundles vs. single journals in discipline
- Library budget numbers/trends
- Real revenue resulting from “long-tail” citation articles
- Percentage of articles covered by federal funds

All of these market conditions regularly contribute to journal cancellations and must be accounted for so that the effect of the embargo period can be adequately isolated. Even if this were possible, it is not clear that different embargoes ought to be considered, even if they imply differential economic effects. For one thing, the effects on subscription-based publishers should be balanced by the economic benefits of widespread access to the economy as a whole. For another, there are substantial costs in increased administration complexity and user confusion as soon as one allows differential embargoes.

## Conclusion

Once again, SPARC deeply appreciates the time and effort that the Office of Science and Technology Policy (OSTP) has taken to lead this discussion on the important topic of ensuring public access to the results of federally funded research.

---

42 Houghton, Rassmusen and Sheehan, page 8. 2010

43 <http://www.roarmap.org>

44 <http://highwire.stanford.edu/lists/freeart.dtl>

COMMENTS - Public Access to Federally Funded Scholarly Publications  
SPARC (the Scholarly Publishing and Academic Resources Coalition) [www.arl.org/sparc](http://www.arl.org/sparc)  
January 12, 2012

SPARC fully supports the expeditious expansion of the current NIH public access policy to all other federal agencies that conduct scientific research, in order to create a freely accessible, permanent digital archive of the results of our nation's investment in scientific research. Our members look forward to providing any additional information that might be useful to OSTP, and to participating constructively as this process moves forward.

To discuss in deeper detail, please feel free to contact Heather Joseph, Executive Director through [heather@arl.org](mailto:heather@arl.org) or (202) 296-2296.

Respectfully submitted,

Heather Joseph  
Executive Director, SPARC

David Carlson  
Dean of Library Affairs, Southern Illinois University at Carbondale  
Chair, SPARC Steering Committee