



Public access to federally funded research: Comments

SUBMITTED TO THE OFFICE OF SCIENCE AND TECHNOLOGY POLICY, JANUARY 19, 2010

FOR MORE INFORMATION, CONTACT HEATHER@ARL.ORG

Introduction

SPARC (the Scholarly Publishing and Academic Resources Coalition), along with SPARC Europe and SPARC Japan, 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 faster and wider sharing of outputs of the research process increases the impact of research, fuels the advancement of knowledge, and increases the return on research investments.

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.¹

SPARC thanks the Office of Science and Technology Policy for convening a robust, open discussion on the importance of ensuring broad public access to the results of federally funded research. We share the Administration's view that enhancing access to this information will promote advances in science and technology, encourage innovation and discovery, and enhance the diffusion of knowledge throughout our society.

We fully support the expansion of the current National Institutes of Health (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.

Why Public Access?

The U.S government funds tens of billions of dollars in basic and applied research each year, with the goals of speeding the pace of scientific discovery, fueling innovation, and – ultimately – improving the public good. Because U.S. taxpayers underwrite this research, they have a right to expect that its dissemination and use will be maximized, and – in particular – that they themselves will have ready access to it. With the recent investment of more than \$15 billion in additional funding for scientific

¹ For more information about SPARC, its members, and its steering committee, visit <http://www.arl.org/sparc>

research provided by the American Recovery and Reinvestment Act,² it is more important than ever to ensure that the investment of this money, and the subsequent reporting on the return on its investment, is as transparent and accessible as possible to the American public.

While the results of this research appear in many forms, articles that report on the results of federally funded research in academic journals are a particularly important subset. With nearly 80,000 papers resulting each year from NIH funding alone,³ these articles represent a significant potential resource for the American public.

However, the reality today is that these articles are not readily accessible to the broad community of stakeholders who may want to read and use them. Currently, the only way to access these articles is to purchase them through subscription or pay-per-view programs run by publishers. For many, access is prohibited by the cost of such programs. Journal subscriptions can cost thousands of dollars each year, with some reaching upward of \$20,000,⁴ and it is common for access to a single article to cost upwards of \$30 through pay-per-view programs.

Networked technology presents us with an unprecedented opportunity to reduce (or eliminate) these access barriers. The Internet can deliver research information today at little to no marginal cost to every researcher, student, teacher, entrepreneur, health care worker, farmer, business owner or any other member of public who may have an interest in reading and applying it in their work and daily lives.

In addition to ensuring rapid and widespread access to this information, we must ensure that these articles can also be more fully used in the digital environment. Simply enabling researchers to read an article is not enough. Given the enormous growth in the number of research articles and terabytes of associated data that are generated each year, researchers need to be able to apply new tools to fully unlock the value contained in this information. They need to be able to use new technologies to identify connections and to make links to articles – and to digital data supporting the articles – in order to enable new research paths and discoveries. They need to be able to data- and text-mine to uncover new contextual relationships that might exist among seemingly disparate elements in these articles.

The full value of our nations' collective investment in scientific research can only be realized if we allow these articles to be freely accessed, used, and built upon. Around the world, there has been a growing recognition that increasing access to research results has great potential to increase the social and economic return on investment in that research. As a result, national agencies, private funders, and discipline-specific research funders are increasingly implementing public access policies. They commonly cite the increased opportunities for new business development, faster R&D growth, enhancement of national research assessment programs, and ensuring competitiveness in the global research community that such policies create as drivers for implementing them.

² See <http://www.recovery.gov>

³ National Institutes of Health Web site (<http://publicaccess.nih.gov/FAQ.htm#f4>).

⁴ The 2008 annual subscription price for the journal *Brain Research* was \$21,744. (<http://www.arl.org/sparc/students> “*The Right to Research*”)

The U.S. should take full advantage of this opportunity, as well. All agencies that conduct scientific research on behalf of the public should be required to ensure that articles reporting on results of all non-classified research they conduct be made available via the Internet for the public to access and to use in a timely manner.

This is particularly urgent in scientific disciplines where new research in biomedicine, environmental sciences, agriculture, energy, economics, education and similar areas can have an immediate and dramatic impact on the public good.

One criteria for determining which agencies are good candidates to enact public access requirements could be the level of scientific research that they fund on an annual basis. Agencies (or departments) with extramural research budgets that exceed a certain amount could be required to enact such policies. SPARC is supportive of legislation currently before the Senate (the Federal Research Public Access Act of 2009, S.1373), which suggests \$100m or more as a benchmark. S.1373 proposes that eleven federal agencies implement such policies, including: Department of Agriculture, Department of Commerce, Department of Defense, Department of Education, Department of Energy, Department of Health and Human Services (including the CDC), Department of Transportation, the EPA, NASA, USGS and the NSF.

SPARC's responses to specific questions raised in the RFI follow.

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?

The process of formally communicating scientific research results through peer-reviewed papers has been a long-standing practice in the scholarly community. The process depends on collaboration among stakeholders in the academy at each step and generally tracks the following path:

- Research is initiated (by federal government, private funders, universities)
- Research is conducted (researchers/authors)
- Results are communicated through journal article authorship (researchers/authors)
- Validity of the research is certified by peer review (conducted by researchers supported by universities or federal government, often organized by publishers)
- Articles are edited, formatted, tagged, and organized for publication (publishers)
- Results are disseminated (by publishers, but also increasingly by authors, universities, federal government)
- Results are archived/preserved (libraries, universities, federal government)⁵

The roles associated with this process have been undergoing significant changes as technology has evolved. This is particularly true for publishing activities such as the distribution of articles, which was simply too expensive for individuals to consider doing in a print-based world. However, the Internet now allows anyone with access to a computer to do these things easily, quickly and at an extremely low marginal cost –

⁵ A good summary of these functions, along with data as to the total annual cost of each function and information about who currently supports the costs, is available in the Research Information Network's report, "Activities, Costs and Funding Flows in the Scholarly Communications System" (<http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu>).

providing an opportunity to reduce some of the friction in the scholarly communication system. This has not gone unnoticed by industry analysts, who have noted the imperative for these types of changes – which can result in economic efficiencies – to continue:

“We would expect governments (and taxpayers) to examine the fact that they are essentially funding the same purchase three times: governments and taxpayers fund most academic research, pay the salaries of the academics who undertake the peer review process and fund the libraries that buy the output, without receiving a penny in exchange from the publishers for producing and reviewing the content... We do not see this as sustainable in the long term, given pressure on university and government budgets.”⁶

Scientists conduct (and the government funds) research so new ideas can be generated, discoveries can be uncovered, and our collective understanding of the world and our interactions with it can be enhanced. Scientists consider the communication of the results of their research to be an essential, inextricable component of doing this work. It is only through sharing results that scientific inquiry moves forward. Research funders and universities share this core mission – the advancement and conveyance of knowledge. SPARC believes that while some of the details of stakeholder roles may change under a comprehensive public access policy, the essential structure of the research process will remain sound.

Scientists and scholars regularly self-organize into communities of practice and expertise for the purposes of peer review and credentialing. These are a part of the culture of science, and will continue. Peer review, in particular, should be allowed to evolve as technology provides avenues for improvement and innovation in how it might be conducted.

Organizations such as publishers, libraries, and aggregators whose operations depend on the results of federally funded research will rightly need to adapt and innovate as federal access policies are implemented and evolve. As with any significant systemic change, some stakeholders might find the demand for the service that they currently provide reduced, and evolve accordingly.

These organizations may also find that this change increases their opportunity to provide new services and products, as was the case when scholarly publishing made the transition from print to electronic formats. A public access policy that enables greater accessibility and usability of research articles will provide broad new opportunities for innovation. With more information available in digital repositories, there will be a growing call for services that locate, organize, and correlate information (among many other things) for end users. The expertise that publishers and librarians have cultivated will be in high demand for the creation and deployment of such new services.

⁶ Credit Suisse First Boston, *Sector Review: Scientific, Technical and Medical Publishing*. April 6, 2004.

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 successful public access policies that have been successfully implemented around the world.⁷ These have been developed after extensive public consultations, and share the goal of maximizing the opportunities that the Internet provides us for conducting and communicating research in a manner that balances the needs of all stakeholders.

The framework that has emerged within these policies is simple and straightforward, and includes a number of homogeneous characteristics that should be included in any policy that the U.S. government adopts:

- Permanent, interoperable, digital archives should be created or leveraged to house digital research results.
- All articles that result (in whole or in part) from federally funded research and have been accepted for publication in a peer-reviewed journal should be required to be deposited into these archives.
- The version of the article may be either the author's final manuscript or the final published version, if permitted by the copyright holder.
- The articles must be made freely accessible on the Internet as soon as practical, but not later than six months after publication in a journal. (The majority of policies currently in use employ a six-month embargo period.)
- The archive must provide for public search and retrieval and full use rights (such as data mining) to these articles.
- Funded researchers must ensure that they retain sufficient rights to comply with the policy when signing publication agreements.
- Compliance with the policy should be a consideration for future funding for researchers.
- The archive should be utilized by the agency to more efficiently manage and understand its research portfolios, monitor scientific productivity, and help set research priorities.
- Funders must provide the public with regular reporting on the extent of author compliance with the policy, and consider non-compliance by funding recipients as a factor in evaluating further grants.

These components reflect a careful effort to ensure that the interests of those who contribute to the funding, creation, and utility of this information be balanced with those of the intended end-users – and ultimate beneficiaries – of publicly funded research, and should be a part of any public access policy implemented by the U.S. government.

⁷ See the ROARMAP (Registry of Open Access Repository Material Archiving Policies) at <http://www.eprints.org/openaccess/policysignup/>

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?

Currently, the primary way to access articles reporting on the results of federally funded research is to subscribe to the scientific journals in which they appear. Libraries at academic institutions represent the largest subscription base for these journals. But, because there are thousands of journals – many of which cost thousands of dollars per year to subscribe to – no institution, no matter how well-funded, can afford to subscribe to all of the publications their campus patrons request access to.

Access to research information is a particular problem at smaller universities, liberal arts colleges, and community colleges, where budget constraints make subscribing to any significant number of journals an impossibility. Yet, there is growing recognition of the importance of the role that these institutions play in preparing an educated workforce for today's knowledge economy, evidenced in part by continued large enrollment increases.⁸ Providing access to even just the subset of articles that result from federal funds would represent a dramatic improvement in access for these institutions.

Members of the public (including physicians, business owners, entrepreneurs, teachers, and others) who are not affiliated with an academic institution must either try and subscribe to journals or purchase articles on an individual basis, which often costs more than \$30 per article. This represents a significant barrier to access for many individuals as well as organizations. The implementation of a federal public access policy would provide a means to open up this layer of information in a timely fashion to those constituencies who currently find the costs prohibitive.

In biomedicine and other health-related disciplines, we have seen a real demand from the public for accurate, up-to-date information. In a recent study, the Pew Research Center noted that 113 million Americans accessed information about health related issues via the Internet in one year. Just as critically, more than half then took that information to a health care provider for further discussion and action.⁹ Though some have argued that peer-reviewed articles are not appropriate for use by the general public, this behavior demonstrates a high rate of willingness to enlist expert help in interpreting and applying this information.

There is also an unmet demand for articles in all disciplines by entrepreneurs and small business owners, who depend on cutting-edge information to help them innovate and build competitive services and products. As an example, a 2009 U.K survey of small and medium enterprises (SME's) by the Publishing Research Consortium (PRC) reported that 73% say they have difficulty getting access to peer-reviewed journal articles.¹⁰

⁸ See, for example, "The Community College Enrollment Boom," *Inside Higher Ed*, at <http://www.insidehighered.com/news/2008/08/22/growth>

⁹ See "Online Health Search," <http://www.pewinternet.org/Reports/2006/Online-Health-Search-2006.aspx>

¹⁰ See page 13, "Access by UK small and medium-sized enterprises to professional and academic information," Mark Ware Consulting LTD, August 2009

In interviews conducted along with the survey, a pharmaceutical company described how it might find 60 to 100 articles of potential relevance to a small project from a PubMed search. Its client would not be prepared to pay \$30 each for all these (i.e. \$1800 to 3000), and the company could not afford to absorb the expense, so they had to select a few to purchase, try to find copies via other routes, or do without. The company could not have known what they missed by being limited in this way, but expressed their belief that it resulted in a lower value in service.¹¹ In an age when we are racing to develop alternative energy sources, green technologies, and sustainable agriculture practices, it is almost unthinkable that we would continue to inhibit access to the results of publicly funded research.

Active researchers and scientists are not immune from the effect of these access barriers. Dr. Sophia Colamarino, a neuroscience researcher whose area of concentration is adult neural stem cells and brain development and regeneration, recounted her recent experience at the annual meeting of the Genetic Alliance in July 2009.¹²

Dr. Colamarino, now VP of Research for Autism Speaks, the nation's largest Autism advocacy and research funding organization, noted that – upon accepting a new position as Science Director for Cure Autism Now – she was dismayed to find herself cut off from the robust access to research journals she had enjoyed while conducting research at the Salk Institute. She noted, “On Friday, I worked at Salk, and had access to this material. On Monday, just because I changed positions I literally found myself cut off.” It was a rude awakening, but not an uncommon one.

She further noted, “As science expands into other sectors of our society there are now many of us who do research jobs without being in a research center, and we are potentially the ones who are most affected by lack of access. Being able to function within the science research community IS my job (e.g. keeping abreast of the literature, being conversant about the newest research advances, and making funding decisions based on this knowledge); yet I don't have access to the material I need to do it.” Graduate students often note the difficulty they experience in trying to stay abreast of recent developments in their fields upon leaving school – realizing that once they graduate their library card expires and, with it, their access to much of the recent literature.

Under a public access policy, articles reporting on federally funded research would be available in digital form in searchable repositories, so that they may also be *used* in entirely new ways. The ability to link, combine, text-mine, or even data-mine these new databases creates rich new opportunities for connections and discoveries to be made – and to potentially be made by an entirely new set of users.

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?

Because of the cumulative nature of research – it gains in value only when used – the value of our collective investment in research can only be maximized by sharing and using findings as broadly as possible. In a 2005 report on scientific publishing, the International Organization for Economic Cooperation and Development put it quite succinctly:

¹¹ Ibid, page 17

¹² See <http://www.geneticalliance.org/conference09>

“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 social returns on public investments.”

Establishing a policy that ensures fast, digital access to the papers reporting on the results of all federally funded research is a critical first step in making this a reality.

To effectively measure the return on our investment in research, both the costs and benefits must be accounted for. The current NIH Public Access Policy provides one example of what these costs might reasonably look like.

The NIH has estimated that its costs to run PubMed Central (PMC) at 100% capacity are ~\$4.5 million year – an amount dwarfed by its annual \$30 *billion* operating budget.¹³ This means NIH leverages a tiny fraction of its budget to ensure broad public access to all of the published results of its funded research. NIH’s investment in this resource has created an opportunity for other agencies to adopt, incorporate elements of, or learn from the NIH’s implementation of their Public Access Policy. To put this into perspective, compare this \$4.5 million to the ~\$30 million NIH spends each year on publication fees and other page charges, which are paid out of grants to investigators.¹⁴

Different agencies will have different levels of article output. With 80,000 articles/year, NIH generates the largest amount of output in terms of research articles published as a result of agency funding. Agencies with a smaller aggregate output may want to consider alternatives to establishing their own repositories. These alternatives might include:

- Using PubMed Central as their repository
- Partnering with other agencies to achieve economic efficiencies in creating a shared repository
- Using the freely available source code for PubMed Central to establish a repository, and save on the costs NIH has incurred to develop and refine it; or
- Partnering with one of the over 1,400 academic and research institutions that have already established robust online digital repositories for scholarly and research content.¹⁵

While it is important for agencies to have flexibility in choosing the location as well as the management strategy for their digital repositories, in order to control systemic costs, it is crucial that policy requirements be closely coordinated to ensure that academic institutions (which will largely bear the responsibility of ensuring compliance) are not overburdened by multiple policies with multiple implementation requirements.

¹³ See “Analysis of comments and implementation of the NIH Public Access Policy,” http://publicaccess.nih.gov/analysis_of_comments_nih_public_access_policy.pdf

¹⁴ See “Policy on Enhancing Public Access to Archived Publications Resulting from NIH-Funded Research,” footnote #8. <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html>

¹⁵ See The Directory of Open Access Repositories, at <http://www.opendoar.org>

On the benefit side, the potential return on an agency's investment in ensuring broad public access can be measured in both social and economic benefits. Models and metrics for gauging the relative return on opening up access to the results of national R&D investments are currently in use, and should be considered. Examples include:

- Houghton, J.W. Steele, C. and Sheehan, P.J. (2006). *Research Communication Costs in Australia, Emerging Opportunities and Benefits*, Department of Education, Science and Training, Canberra.
- Department of Trade 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/EI-ASPM>)

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

Experience from public and private funders who have successfully implemented public access policies has shown that, in order for a policy to fully succeed, it must – first and foremost – be a requirement, not a recommendation. The NIH has published statistics that quite compellingly illustrate their experience with low researcher compliance under a voluntary public access program, and the subsequent dramatic increase once the policy became mandatory.¹⁶

Ensuring researcher awareness is a key issue. Funders, working in close collaboration with academic institutions, have had good success in increasing compliance rates when they provide regular, clear messages about expectations through communications channels to which researchers currently pay attention – including campus grant networks, reminder notices for regular progress reports, grant summaries, year-end reports, and similar venues.¹⁷

Perhaps just as important as requiring compliance is making it as easy as possible for researchers to comply with a policy. The mechanisms that researchers are asked to use to comply should be as closely integrated into their normal workflow as possible. For example, article submission formats should reflect what is in broad use in the community. If possible, a single, common user interface should be presented to users when they are asked to deposit manuscripts into federal repositories – for example.

To have the greatest chance of success, federal public access policy requirements should be standard across all agencies. Because the vast majority of federal research recipients are actually academic institutions, they will bear the primary responsibility for ensuring compliance. Offices of sponsored research, grant offices, and academic departments should not be overburdened by multiple policies with multiple implementation requirements.

¹⁶ See <http://www.nihms.nih.gov/stats/>

¹⁷ See “Implementing the NIH Public Access Policy is as Easy as A, B, C,” from the University of Texas at Austin, for an example - http://www.utexas.edu/research/osp/open_access/index.html

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)?

Access to the final published version of an article is preferable, so long as full use of the article is permitted. If this is not possible, access to the author's peer-reviewed manuscript is an acceptable substitute. In this case, articles should be clearly identified as author manuscripts, and should include a link to the final published article. Agencies should also follow the practice established by NIH and other funders of encouraging publishers to replace the author's peer-reviewed manuscript with the published edition if they choose to.

Requiring access to the author's final manuscript is a common practice for research funders with public access policies.¹⁸ The practice evolved as a mechanism to protect subscriptions and help journal publishers earn a return on their investment in the copyediting, formatting and other value-added services that they may put into the final published version.

While some have raised concerns that the author's manuscript may contain errors or omissions that are only corrected during the copyediting process, the reality is that many journal publishers already routinely provide access to the author's manuscript on their own Web site prior to the journal's publication. (Such programs are called "Publish Ahead of Print" or "Papers in Press"). These programs have been in routine use for years, and are touted as a benefit to scholarly society members as well as to the wider scientific community. Some examples of publishers in a wide variety of disciplines who do this include:

- The American Diabetes Association
(<http://care.diabetesjournals.org/papbyrecent.shtml>)
- The American Geophysical Union
(<http://www.agu.org/contents/journals/ViewPapersInPress.do?journalCode=GL>)
- The American Speech-Language-Hearing Association
(<http://jslhr.asha.org/papbyrecent.dtl>)
- The Society for Environmental Toxicology and Chemistry
(<http://www.setacjournals.org/perlserv/?request=get-toc-aop&issn=1552-8618>)

As scholarly communication practices and norms evolve in a digital environment, the definition and importance of what constitutes a final article will evolve, as well. Many disciplines are already experimenting with post-publication comments and peer review practices, making the publication process more of an ongoing, rather than a periodic, one. The standards used to connect these comments, as well as subsequent alterations to an article, will provide further mechanisms to help scientists and the public effectively navigate versioning issues.

¹⁸ See, for examples, <http://www.sherpa.ac.uk/juliet/>

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?

Ideally, peer-reviewed research papers should be made publicly available as soon as they are published, to ensure maximum impact and maximum improvements to the public good. However, recognizing that immediate public access could have a negative impact on non-open access publishers that are dependent on subscription revenue, an embargo period of up to six months is an acceptable compromise with the public interest.

The idea that embargo periods might vary in response to the version of the articles provided is a potentially very interesting mechanism that might be used to ensure that ultimately – after an embargo period of no more than a year, for example – articles under a public access policy are truly “Open Access.” True Open Access is to say that articles can not only be read, but also downloaded, searched, linked, mined etc. as called for by the Budapest Open Access Initiative.¹⁹

An embargo period of six months or less is the norm in public access policies being implemented in a growing number of countries worldwide, by both private and public funders. Some notable examples include: The Australian Research Council, Arthritis Research Campaign, British Heart Foundation, Canadian Institutes for Health Research, Canadian Breast Cancer Research Alliance, Canadian Cancer Society, Department of Health UK, The European Research Council, the Austrian Science Fund, Genome Canada, The Higher Education Authority (Ireland), JISC (UK), Principo de Asturias (Spain), the Science Foundation of Ireland, and the Swedish Research Council.²⁰

Six-month embargo periods are also currently in use by journal publishers. All 90 journals published by the Nature Publishing Group in a variety of disciplines (many in partnership with scholarly societies) have a policy permitting a six-month embargo period.²¹ These include journals in a broad range of disciplines, including Nanotechnology, Physics, Photonics, Materials, and Chemistry.

Dozens of other journals with a six-month or shorter embargo period are listed on the Web site of HighWire Press.²² Titles with a six-month or shorter embargo include large, well-known publications such as the *Journal of the American Medical Association*, *Proceedings of the National Academy of Sciences*, and the *New England Journal of Medicine*, as well as more discipline-specific, society-based publications such as the *Journal of Neuroscience*, the *American Journal of Pathology*, *Diabetes Care*, and others.

The primary concern for subscription-dependent publishers is cancellation by libraries, which represent roughly 80% of their revenue. However, to date, journal

¹⁹ Online at <http://www.soros.org/openaccess/read.shtml>

²⁰ Also detailed at <http://www.sherpa.ac.uk/juliet/>

²¹ See the NPG author license policy at http://www.nature.com/authors/editorial_policies/license.html

²² At <http://highwire.stanford.edu/lists/freeart.dtl>

publishers have reported no evidence to suggest that policies with embargo periods of six months or less have a negative effect on the health of their publishing programs.

Additionally, predictive studies, including a report by a prominent publishing trade group, the Association of Learned and Professional Society Publishers (ALPSP), have outlined the conditions that would have to exist in order for libraries to cancel journal subscriptions, as related to availability of material in a publicly accessible archive:

1. First, an extremely short embargo period is necessary. 82% of librarians surveyed noted the embargo period would need to be three months or less before they would consider it a factor in cancellation decisions.
2. Second, the final publisher's version would need to be available. Librarians reported that the raw manuscript, or preprint, is not a substitute for the journal. And, only 9% saw access to an author's final manuscript as an adequate substitute for the final manuscript.
3. Third, comprehensiveness counts; 75% of librarians indicated the archive would have to contain over 90% of a given journal's content before it became a factor in possible cancellation.²³

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?

To facilitate the widest possible compliance, standards for archiving and interoperability need to be enacted on multiple levels – the technical, repository, article and data/article component levels.

On the article level, submission formats should reflect what is in broad use in the community (i.e., common formats like Word, LaTeX, etc.). However, to ensure that articles are ultimately in a format that allows rich searching, linking, text mining, etc. they should be converted to a standardized mark-up language designed to facilitate these functions. XML is the current standard in scientific publishing.²⁴

To further enhance the usability of the information contained in these articles, a common, standard document type definition (DTD) should be adopted. The National Library of Medicine (NLM) DTD is currently in broad use in the scholarly publishing community – by stakeholders ranging from the Library of Congress and British Library to the HighWire Press – and should be given serious consideration as a standard.²⁵

The evolution of Digital Object Identifiers (DOIs) has created the capacity to ensure standards for searching and linking at an even more granular level. DOIs have become an industry-accepted standard mechanism to facilitate the persistent identification, location, management, and linking of electronic articles and manuscripts, as well for

²³ Summary and Conclusions of the study may be found at http://www.alp.org/ngen_public/article.asp?id=200&did=47&aid=157&st=&oaid=-1

²⁴ See <http://www.w3.org/XML/>

²⁵ Online at <http://dtd.nlm.nih.gov/>

components of these articles (images, graphs, etc.) DOI names can be used for any form of management of any data, whether commercial or non-commercial. The system is managed by the International DOI Foundation, an open membership consortium including both commercial and non-commercial partners that has recently been accepted for standardization within ISO.²⁶

Standards for digital repositories to interoperate also must be enacted. Currently, the Open Archives Initiative (OAI) offers an established and successful set of standards for metadata harvesting (OAI-MH) as well as object reuse and exchange (OAI-ORE). These should be considered as a basis for any standards adopted by federal repositories.²⁷

Another set of important considerations in ensuring true interoperability are legal standards. The terms of use that are associated with articles are just as critical to their downstream utility as any technical component. Authors and end users must have the rights to make full use of these articles in order for them to achieve their full value. Creative Commons offers a set of licenses that should be considered to standardized terms of use.²⁸

As with any standards, these can – and should – evolve over time. A key strength of all of the standards proposed here is that they are open standards, which are widely supported by the user community, and are regularly updated in direct response to the community's demonstrated needs.

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?

This question speaks to the heart of the motivation to enact a public access policy. We have entered a new era in which the conduct and reporting of science is fully digital. The amount of data that is collected during the research process is growing exponentially; even the small percentage of data represented by peer-reviewed digital articles is vast. It's impossible to expect that a person – or even a group of people – will be able to fully process the mountains of new information created each day in a meaningful, complete way. Computers are now just as important a category of end-users of this material as people are.

As we realize the increasing value of interdisciplinary research, the ability to move from a word or gene or chart or micrograph noted in an article in one discipline to a word or structure in a seemingly disparate discipline based on textual or semantic clues that only advanced computational programs can provide becomes critical. The ability to digest huge amounts of information and uncovering broad trends in mountains of scientific articles that are no longer locked away holds tremendous promise – as does the possibility of parsing these articles into their smallest possible components, and finding connections among their tiniest elements. This kind of work is only enabled if these articles are available in an open, standard and accessible manner.

²⁶ See <http://www.doi.org/>

²⁷ See <http://www.openarchives.org/>

²⁸ See <http://www.creativecommons.org>

As a result of the NIH Public Access Policy, biomedical researchers can move from references to gene sequences embedded in the text of an article to the gene sequences themselves in the Genbank database. In this same way, researchers at NOAA or the EPA should be enabled to make links from research articles to data sets on climate change, for example, and to apply data mining techniques, modeling programs, and the like to accelerate understanding in crucial arenas that directly impact the public good.

This means ensuring not only that anyone can access articles to read at anytime, from any place, but also that they have the ability and the rights required to use them in new and innovative ways. The rights of end users to read, download, copy, link, crawl, data-mine, or reuse for any other lawful purpose – must be clearly spelled out as part of any federal policy. As noted above, Creative Commons provides licenses that enable such reuse rights, and should be considered.

We can only begin to imagine the possible uses and applications that the scientific community – and the public – might envision and construct to truly unlock the value of the results of federally funded research. A comprehensive policy to ensure that they have the ability to begin to do so is a goal that we should collectively work towards making a reality as soon as humanly possible.

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.

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 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

Ray English
Azariah Smith Root Director of Libraries, Oberlin College
Chair Emeritus, SPARC Steering Committee