

Boston College University Libraries

Response to:

OFFICE OF SCIENCE AND TECHNOLOGY POLICY

Request for Information: Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research, November 3, 2011

From:

Thomas B. Wall, University Librarian

The Scholarly Communication Committee

Boston College University Libraries

Chestnut Hill, MA

(1) Are there steps that agencies could take to grow existing and new markets related to the access and analysis of peer-reviewed publications that result from federally funded scientific research? How can policies for archiving publications and making them publically accessible be used to grow the economy and improve the productivity of the scientific enterprise? What are the relative costs and benefits of such policies? What type of access to these publications is required to maximize U.S. economic growth and improve the productivity of the American scientific enterprise?

Key points

- Full and immediate open access to products of taxpayer-funded research has the greatest potential for:
 - Rapid spread of ideas
 - Increased commercial benefits
 - Acceleration of scientific discovery
 - Increased interdisciplinary discovery
 - Enabling the unintended or disadvantaged reader
 - Enabling computer analysis
- Existing business models can adapt, with help from grant-funders during the transition.
- Additional costs would be small, for great benefit.

Full immediate open access to the products of taxpayer funded research holds the most promise for rapid spread of ideas. This strategy has the potential to create the greatest return on investment for taxpayer money and to increase commercial benefits. Access should be online, without economic barriers, free of technical barriers and coupled with the right to reuse and build upon the knowledge accessed. It is important that users have the ability to reuse, redistribute, make derivative works, datamine and reconfigure content as necessary to maximize potential benefit and leverage data to innovate.

The economic well-being of the nation benefits from open access as the government's return on investment in research is augmented. Access to previously hard to obtain research encourages innovation, stimulates new ideas, services, products, and creates new opportunities for job creation. Research is exploited

through new and diverse pathways that lead to greater modernization and commercialization of many aspects of the US economy. Open access to the results of federally funded research is beneficial to other nations too, particularly less developed ones. Institutions and people in developing countries find it exceedingly difficult to afford the scholarly journals published in the developed world. Having open access to such publications provides greater opportunities to researchers in developing nations to engage in better scientific, medical, technological, business and other research. The American public has the opportunity to become better educated so that it can more meaningfully understand and support scientific, industrial and related enterprises. Open access to the results of federally funded research provides them with previously unavailable materials relating to health, energy, environment, and other areas important to day to day living. Similar concerns and goals were expressed recently in the UK government report, *Innovation and Research Strategy for Growth*, which stated:

The Government, in line with our overarching commitment to transparency and open data, is committed to ensuring that publicly-funded research should be accessible free of charge. Free and open access to taxpayer-funded research offers significant social and economic benefits by spreading knowledge, raising the prestige of UK research and encouraging technology transfer. At the moment, such research is often difficult to find and expensive to access. This can defeat the original purpose of taxpayer-funded academic research and limits understanding and innovation... .[W]e need to go much further if, as a nation, we are to gain the full potential benefits of publicly-funded research.

Innovation and Research Strategy for Growth, Presented to Parliament by the Secretary of State for Business, Innovation and Skills, December 2011. <http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf>

Currently, university graduates who have become accustomed to full access to the peer-reviewed literature in their field find themselves, as practitioners in the field, with no access to the literature. Their access to new developments reported in the broad range of scholarly journals will be frozen at graduation. Requiring that taxpayer-funded literature be made openly accessible to all will allow current practitioners in all fields to have equal access to the literature. Making the literature open to all in a free database will give the high school student, the inventor in the garage, the small businessman and the clinician in a non-profit clinic the same access to new knowledge as the scientist with a large R&D budget at his disposal. Discovery should be made easy by allowing indexing and searching by common search engines as well as by the database's own interface.

When publications are produced in the traditional manner, much time can elapse between the end of the research phase and the final publication. Many disciplines have found the length of this cycle prohibitive to their advancement. In the field of physics for example, the arXiv preprint server fills a perceived need to disseminate research more quickly. This wider, faster distribution can accelerate the process of scientific discovery and innovation.

The increasing interdisciplinarity of research and approaches to complex scientific problems would also argue for broader access. Even an enterprise with a large R&D budget will likely put its money only into the publications most central to its field. With the costs of publication rising at a rate many times that of the CPI, subscription cuts have been necessary for many. (See *Monograph and Serial Expenditures in ARL Libraries, 1986-2004*. <http://www.arl.org/bm~doc/monser06.pdf>)

It is not hard to imagine that answers to climate change issues might be found in research in biology, physics, material science or mathematics, yet many scientists now have access to a narrow range of publications central only to their field. Open access to taxpayer funded research would eliminate economic barriers to knowledge and maximize exposure of the research to foster scientific productivity in any quarter. Just as wider access to the literature associated with the Human Genome Project brought unprecedented advances in scientific knowledge, unfettered access could again result in major strides toward addressing this looming global threat.

Open access also provides an opportunity for a new kind of reader with new capabilities. If the contents of the articles (and the underlying data) are openly accessible in compliance with metadata standards, they can be “read”, digested, analyzed, and compared by computer programs that recognize new patterns and derive new knowledge based on massive amounts of data. If the articles are closed and inaccessible, this is not possible. This process can help scientists identify relevant data more efficiently and accurately. It opens avenues for IT innovators to create tools and processes to help with data analysis, strengthening and providing new areas of growth for the IT industry.

By far the greatest cost of funding research and making it openly accessible is the cost of the research itself. Taking the NIH experience as a model, the cost of making the research available in an open access database is only .01 of the overall NIH budget. That incremental increase in cost makes 2.2 million articles available for download by 500,000 persons per day. (See testimony of David J. Lipman, M.D., Director, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, U.S. Department of Health and Human Services (HHS) on Public Access to Federally-Funded Research before Committee on Oversight and Governmental Reform, Subcommittee on Information Policy, Census and National Archives, United States House of Representatives.

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

Many of these articles are published traditionally as well. The 12 month embargo period that NIH allows provides an economic balance – allowing the publishers to make their initial profits (recouping their costs during the time) but making the information accessible fairly soon. This embargo period might be shortened or eliminated by providing other financial incentives to those publishers who publish articles produced with federal funding. The article publishing fees that many publishers charge could be funded through the author’s grant, providing a substitute revenue stream for publishers.

An important analysis of the potential benefits to result from making federally funded research open access is the July 2010 report to SPARC [The Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs](#) (John Houghton, Bruce Rasmussen and Peter Sheehan). The primary focus of this study was the FRPAA bill. The report states:

Over a transitional period of 30 years from implementation, the potential *incremental* benefits of an open access archiving mandate for all FRPAA agencies’ funded R&D[2] might be worth around \$1.6 billion (Net Present Value), around 4 times the estimated cost using the higher end lifecycle costing, 8 times the cost using NIH costing and more than 24 times the cost using arXiv costing. Perhaps some \$1 billion of these benefits would accrue within the US, with the remainder spilling over to other countries. Hence, the US national benefits might be around 5 times the costs.[3] The overall impacts of openly archiving all FRPAA agencies’ funded R&D article outputs would be greater than these incremental impacts, with likely US national benefits of around 8 times the costs (Table 4). These estimates assume a six-month embargo period between publication and open accessibility. If there were no embargo, we estimate that incremental returns might be closer to \$1.75 billion.

This data indicates that immediate open access would produce the greatest returns.

(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? Conversely, are there policies that should not be adopted with respect to public access to peer-reviewed scholarly publications so as not to undermine any intellectual property rights of publishers, scientists, Federal agencies, and other stakeholders?

Key points

- Publication services and compensation can evolve.
- Lack of taxpayer access is the primary imbalance to address.
- Open access will be beneficial to authors' careers.
- Open access and standard licensing would eliminate chilling risk calculations by users of content.

Publishers who are not creators have no inherent intellectual property rights. They obtain them from author/creators as compensation for their publishing services. Exploration of new business models can find new ways to compensate publishers for a new array of useful services. It may be that we no longer need print content with high cost and limited distribution. We may still want the editorial, design, and distribution services and networks that publishers provide. Currently, embargo periods serve to allow publishers to recoup costs, but we may be able to pay those costs directly as a very small part of the research funding. Beyond the transitional period, it is possible that traditional publishers will change into providers of a different type of service, one that provides content freely but, as an example, also provides analysis and data mining tools and services for a fee.

As needs change, publishers will evolve to continue to meet those needs, but this should not be done at the expense of public access to publicly funded knowledge, or at the risk of slowing the progress of science to a pace that meets nineteenth century distribution methods. Publishers are already proving that they can innovate in this area, by providing deposit services to authors after embargo periods and by creating hybrid journal models where traditional closed articles coexist with openly readable articles.

Currently, in cases where open access to research is not required, the public is accorded no rights in the research it funds. This is the primary imbalance in the system that should be addressed.

Scientists typically receive little or no monetary compensation for the intellectual property created by writing and peer-reviewing articles. Their compensation comes in the form of enhanced career prospects, recognition and advancement within their fields. Making their articles openly accessible would not adversely affect them and in fact may advance their careers even further as new applications are built on their discoveries. Open access allows greater readership and the citation counts on which their careers are currently are often evaluated. (See Swan, A. (2010) The Open Access citation advantage: Studies and results to date. Technical Report, School of Electronics & Computer Science, University of Southampton.

<http://eprints.ecs.soton.ac.uk/18516/>)

By making all articles openly accessible and by allowing adoption of standard licenses defining reuse we can eliminate ambiguity that currently plagues the intellectual property environment. Merely extending the existing NIH mandate to other agencies would not accomplish this. The current NIH mandate is deficient in this respect. Rights of reuse should be defined and stated upfront. The current chilling effect of the need to evaluate risk would be eliminated and no one would need to fear infringement in their legitimate reuse of scholarly work.

(3) What are the pros and cons of centralized and decentralized approaches to managing public access to peer-reviewed scholarly publications that result from federally funded research in terms of interoperability, search, development of analytic tools, and other scientific and commercial opportunities? Are there reasons why a Federal agency (or agencies) should maintain custody of all published content, and are there ways that the government can ensure long-term stewardship if content is distributed across multiple private sources?

Key points

- A centralized system could apply one set of standards for applications, metadata, access and interoperability.
- Federal operation would ensure longevity and sustainability and facilitate international cooperation.
- Government/academic collaboration would take advantage of collective expertise.
- Geographic distribution of mirrored content would be vital for digital preservation.

A centralized approach would employ one set of standards so that access methods and applications, methods of analysis and data mining would be made easier and more universal. We recommend a collaborative partnership between government and universities for the stewardship of these digital materials. This arrangement would benefit from the collective expertise of its stakeholder-participants, minimize any duplication of effort and maximize interoperability. Many universities have already joined efforts to promote discovery, access, and preservation for their extensive digital collections. And, many of the lessons they have learned in doing so would be relevant to ensuring long-term curation of published articles.

If the centralized system were federally operated, that would ensure longevity and sustainability, without the uncertainty of continued operation of private entities. A national archive would also facilitate international cooperative efforts, which will be extremely important for global scientific advancement. Openly accessible content would, by definition, be better preserved because it could be copied and distributed in geographically diverse preservation archives without running afoul of copyright restrictions. Greater regular use would ensure that any degradation in quality of the digital data would come to light more quickly. Geographic distribution of identical content would be vital for digital preservation. Several preservation initiatives serving libraries, such as PORTICO, LOCKSS and the MetaArchive Cooperative may provide some model components of a preservation program.

Some attributes that would promote trustworthy stewardship include:

- Conformance with the Open Archival Information System reference model
- Multiple copies –widely dispersed geographically and with fixity checking on a regular schedule
- The use of non-commercial open source software
- The use of non-proprietary file formats and a plan for future migration
- Adherence to best practices and standards, especially for metadata creation
- Trustworthy Repositories Audit & Certification for all membership repositories (Currently draft international standard ISO/DIS 16363. For more details, see: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf)
- Mirrored and centralized online searching

(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?

Three existing models come to mind:

- The current relationship between taxpayer-funded research grants awarded by NIH, often to researchers at private institutions, resulting in publication by private publishers and then deposited and made openly accessible (although not necessarily free of copyright constraints) in the public database, PubMed Central.
- The collaborative efforts of the HathiTrust members, consisting of some private and some public institutions, engaged in a cooperative enterprise to build a database of published literature, making it accessible to all. In addition to the nearly 3 million public domain volumes available to all, this project has resulted in nearly 10,000,000 volumes being available and accessible to visually impaired readers.
- MetaArchive Cooperative, a group consisting of public and private university partners, engaged in a cooperative enterprise to provide digital preservation for the member institutions' digital scholarly publications. The diversity of membership allows for shared, duplicated, geographically distributed preservation of content.

(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?

To avoid excessive costs, clear guidelines should be developed to define a minimum core and as much metadata as possible should be populated from dropdown lists and other assistive methods. A schema which is granular enough to support services such as open url generation, creation of derivative reports and products, and inclusion in search engines is important. MODS comes to mind as a schema that would work.

Note that the Google Scholar inclusion guidelines for webmasters indicate that the appropriate metadata schema for journal articles must have unambiguous fields for journal title, volume, issue, and page numbers. Some popular schema, such as Dublin Core, are deficient in this regard.

<http://scholar.google.com/intl/en/scholar/inclusion.html>

(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?

Key points

- Representatives of all stakeholder groups should be involved in setting policy.
- The policy should be simple and consistent across all Federal agencies.
- All agencies should allow use of grant funds to pay article processing fees.
- Publishers should handle deposits to public databases.
- Databases should allow easy discovery and interoperability.

In order to make sure that concerns of all stakeholders are recognized and addressed, representatives of all groups should be involved in making policy.

It will be important to craft a policy that is simple and consistent across all funding agencies. Currently the differing requirements can be frustrating for grant applicants. Grant application requirements should be uniform, except to the extent required by disciplinary differences. If embargo periods are established before public access is allowed, these should be as uniform as possible. Again, differences in disciplines will be important and should take into account legitimate intellectual property interests of scientists and institutions in development of patents.

All agencies should allow use of grant funds to pay article processing fees so that publishers can recoup direct costs. Publishers should be required to handle public deposit, relieving researchers of the need to keep track of embargoes and deposit mechanisms. Deposit track records should be a factor in grant awards, both for researchers and for publishers – and these should be publically available information.

Central deposit would have the benefit of the ability to generate institutional and individual pages tracking content and contributions which could be used in institutional assessments. Standardized deposit mechanisms and metadata would be key to this kind of data reuse. Librarians and IT professionals should be involved in database design to make sure that material in the database is easily discoverable and interoperable with current and developing systems.

(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?

Key points

- All types of publications resulting from funded research should be made available open access.
- Books and conference proceedings may require different business models than journal articles.

Most types of peer-reviewed publications that result from federal funding should be made immediately available as open access. This would certainly include book chapters, articles in conference proceedings and technical reports. Different economic interests may be at stake, however, requiring a somewhat different analysis.

Academic authors are not usually financially compensated for journal articles. Textbook authors may, however, receive substantial royalties. If the foundation of the work is taxpayer funded research, the product of the research should be publicly available, but some portion of the grant award might be used to compensate both publishers and authors, at least in a transitional period. Online publishing, while not free, requires a significantly smaller investment from the publisher. One model for books/monographs that result from federal funding might be for a portion of the grant to be set aside as a subvention for a) publishers to produce the resultant publication and b) as appropriate royalties for the author. Another strategy is to publish with a publisher of open access books. A growing number of important presses are engaged in publishing OA books. See [http://oad.simmons.edu/oadwiki/Publishers of OA books](http://oad.simmons.edu/oadwiki/Publishers%20of%20OA%20books)

Open standards should be developed for resulting e-books. Currently the e-book landscape is highly diverse, unstandardized and difficult for individuals to navigate. Publishers might be provided with incentives for innovative and universal solutions to these issues. The chaotic landscape provides publishers of vision with an opportunity for innovation.

Conference proceedings should also be made available if the research is publicly funded. Consideration should be given to impacts on scholarly societies. Revenue for associations with dwindling memberships is going down. Receipt of conference proceeding volumes is one of the incentives for payment of membership dues. Like publishers, societies may need to investigate providing different types of value and services to their members to maintain their income. If grant funding is allowed to compensate for journal processing fees, it might also subsidize open online conference proceedings.

(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?

Key points

- Access should be immediate if possible.
- The publishers' business models are evolving.
- Disciplinary differences will need consideration.

In terms of maximizing benefits of taxpayer funded research, the current traditional publishing system has faults that are ingrained by its origins in a print culture. Compared to the speed and global dissemination of an internet-based distribution system, it is antiquated and slow.

For maximum benefit to the public, free and immediate open access would deliver the most benefit.

Ideally there should be no embargo period – the results of federally funded research should be available immediately. Publishers should permit author open access self-archiving. However, for the near future it is likely that many publishers which allow some aspect of open access will demand a delayed access period in order to protect their sales. This model assumes that scholars want the most up-to-date articles immediately upon publication, and that educational institutions will continue to pay the subscription fees to access them. Scholars do not want to wait for the end of the embargo period to read journal articles. Thus embargoes, though beneficial for publishers, are problematic for scholars, students, and the public who have often paid for much of the research. They do damage to the principle of genuine open access and delay the benefits of publicly funded research.

The Houghton report stated:

Hence, a six-month embargo reduces the returns by around \$120 million (NPV). Of course, the impact of an embargo delaying open accessibility will vary significantly between fields of research and disciplines, having greater impact in faster moving fields of research and practice than in those where the progress of knowledge, application and practice is slower.

[The Economic and Social Returns on Investment in Open Archiving Publicly Funded Research Outputs](#)

In short, if embargoes have to be implemented, it is desired that the delayed access period be *as brief as possible*. An embargo period of twelve months or less has been voluntarily adopted by many commercial journal publishers. This indicates that they do not think a longer period is necessary for their business model. (See <http://highwire.stanford.edu/lists/freeart.dtl> and

<http://romeo.jiscinvolve.org/wp/2011/11/24/60-of-journals-allow-immediate-archiving-of-peer-reviewed-articles-but-it-gets-much-much-better/>)

Disciplinary differences may need to be adopted. The creation of preprint servers in some fields indicates that the value of information in some disciplines has a shorter half-life. That appears to be the case in physics and computer science (shown by extensive use of arXiv), in economics (RePEc contributions are high) and in law (extensive use of SSRN is common). The development of these repositories appears to be driven by inherent discipline-based need among scholars. It may be that public need, particularly in scientific disciplines where the stakes are high, necessitates consideration of shorter embargo periods.

If publishers are compensated differently for their contribution to the publishing enterprise than by the current subscription system, they would not need to delay access to allow for subscription. To maximize availability of publicly funded research, alternatives to the traditional subscription system must be found. In addition, it should be noted that library budgets have been severely impacted by rising subscription costs. Not only is this trend unsustainable but it is symptomatic of market failure. The users of the information (scholars) are largely insulated from the increased costs. An interesting analysis of this economic imbalance is available in an interview with Prof. Stuart Shieber, Director of the Office for Scholarly Communication at Harvard University.

<http://scientificdatasharing.com/general/interview-with-stuart-shieber/>

Given the conclusions in the Houghton Report, cited above, about the adverse effect of even a six-month embargo, any embargo should be as short as possible. Embargoes should be seen as temporary measures intended to help the publishing industry during a transitional period. It would be helpful to define the end date of that transitional period, by which embargoes would be phased out.