**Request for Information: Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research**

John Wiley & Sons (Wiley) is pleased to respond to OSTP's November 3, 2011 Federal Register notice requesting comments on "Public Access to Peer-Reviewed Scholarly Publications Resulting from Federally Funded Research." We appreciate the opportunity to participate in the Administration's consultation with stakeholders in the scientific research enterprise.

Founded in 1807, Wiley is North America's oldest independent publisher, and has a distinguished history as a literary, scientific, technical, medical, and scholarly publisher, serving researchers and practitioners in the US and around the world. Today, we employ approximately 2,600 staff in the United States and 5,300 globally. We are one of the world's foremost academic and professional publishers. We publish over 1,500 scholarly peer-reviewed journals, and our online service Wiley Online Library (http://onlinelibrary.wiley.com/) provides electronic access to more than 5.5 million articles (as of December 2011) across these journals. Wiley-Blackwell, Wiley's scientific, technical, medical, and scholarly publishing group, is also the world's largest society publisher, working in partnership with over 800 learned and scholarly and professional societies which represent millions of members globally. These include the American Cancer Society (ACS), for which we publish *Cancer*, their flagship journal; the Sigma Theta Tau International Honor Society of Nursing, with more than 120,000 members; and the American Anthropological Association, for which we publish 23 journals. Many of the societies we and other publishers partner with depend to a significant extent on the revenues generated by publishers to support activities which benefit the communities those societies serve and the general societal good.

We publish subscription-based and open access journal titles. For instance, with the The American Heart Association (AHA) we recently announced an innovative venture to publish a new open access journal, Journal of the American Heart Association.  This journal will serve as the first online-only open access journal for the AHA, and joins the AHA's prestigious portfolio of 11 peer-reviewed print and online subscription-based scientific journals.

As preamble to the RFI questions –

- The phrases "the results of research" and "peer reviewed scholarly publications" are used interchangeably in the RFI; they are not the same thing. We agree that the taxpayer should have full access to the results of federally-funded research. Any report which a federal (or any other) research funder chooses to commission from a grantee describing the results of the grantee's research and adds to a publicly-accessible database would be useful in promoting access to the results of research. Neither publishers, nor the learned societies for whom we publish, own the results of research or claim any ownership of any article until it is submitted to us for publication, when we begin to add value through peer review and other activities in the publication process. What we copyright and own is the value-added article; the article becomes our work product. Expropriation of these value-added articles by the government without compensation undermines copyright, intellectual property rights, jobs, and exports. Funding agencies receive regular reports from grantees that can be made publicly available, as the government deems appropriate, likely at minimal additional cost. Why would our government choose to take a publisher's or a society's work product, rather than commission a separate report from grantees? Because it recognizes the value of the publisher value-add but does not want to pay for it. Government funds much scientific research but that does not entitle it to access to and control of the journal articles arising from this research. Taxpayers may fund the research, but they do not fund the publication of this research and, therefore, should have no expectation of receiving free access to this material.

- The RFI also uses the phrase "long-term stewardship" repeatedly, implying that stewardship is an issue. Since the Royal Society launched the Philosophical Transactions in 1665, learned societies and journal publishers have been stewards of the literature. Indeed, since journal content delivery via the web was initiated in the mid-1990s, many publishers have invested not only in the technology necessary to deliver this content online, but also in the digital recovery of print material, often back to the very first issue of a title. This is true for all titles published by Wiley-Blackwell. By one estimate publishers have recovered 40 million articles previously available only in print and made them available online.

- The combination of investments in digital and online technology (by publishers as well as others), and the formation of library consortia (assisted by publishers in many cases) around the country and the world, has accelerated and broadened access to the peer-reviewed literature by orders of magnitude. There is more access to more content by more users now than ever before. Publisher innovation and investment over the past 15 years has made this possible and ensured "broad public access". Wiley Online Library users now have access to 5.5 million journal articles. The platform will be visited by an estimated 100 million unique visitors in 2011, who will download 200 million articles; growth in usage and citations continues to be explosive. Usage of Wiley-Blackwell titles will grow, we project, by 60% in 2011 compared to 2010. Moreover, data also shows that, while academics rank information as 5th out of 15 factors important to their success, access to this information is ranked 13th out of 16 factors as a barrier to success. Most academics and researchers access the necessary literature via subscriptions or licenses maintained by their institutions. ([Access by UK Small and Medium-sized Enterprises to Professional and Academic Information](), Mark Ware Consulting Ltd for Publishing Research Consortium – April 2009).

- Many research funders require research progress reports on all grants. Expanding this information by requiring the addition of a one-paragraph lay summary, and making both freely available, has more potential to enhance public understanding than does providing free access to scientific journals. Our strong preference would be that the federal government does not mandate deposit of journal manuscripts in a freely available archive, regardless of format, process, or timing. Rather, the federal government should strive to provide public access to the information that it already controls and has a right to distribute — for example, research summary reports.

  These reports are produced as part of each federally funded project and they are delivered to the government as a contract deliverable, thus there is a report for virtually every project. The project itself underwent peer review before being selected for funding, and the research results being reported on are solely that which the government funded. In short, these reports are the federally funded research results. Thus if the policy is to provide public access to federally funded research results, then these reports are the natural vehicle for doing so. The government already has them, so all it has to do is make them publicly available. Several federal science agencies already do this; no new system is required.

  A great deal of federally funded research is of an applied nature, as opposed to being basic research. For example, in the Department of Energy applied programs account for a large fraction of the research budget. In many cases, applied research results are not published as journal articles, just in the project research reports. Yet this applied research is often the most suitable for technology transfer into the private sector, which is one of the major goals of public access policy. A focus on journal articles misses much applied research.

  In addition, negative results are seldom published as journal articles, but are often useful and are described in research reports.

- Serious errors in manuscripts are frequently corrected after the peer review process.  For example, in 2010 alone, we issued more than 1,000 corrigenda, errata, statements of concern and retractions with respect to published articles. We are extremely concerned that ~~using~~ making available any version other than the true "final" one (the "Version of Record") will cause confusion, at a minimum, and could significantly compromise the scientific record. We take seriously our role as the stewards of the research literature and version control is an important component of that role.

- Mandating a single approach to public access will stifle innovation in what is now a rapidly changing environment, both by decreasing the amount that publishers are able to invest and by reducing our – and our society partners' - incentive to develop new tools, delivery vehicles, and functionality.

In summary, publishing is a business underpinned by the copyright laws of the US and almost all other countries. However, in the vocabulary of many current anti-copyright activists, "public" is being conflated with "free." There is already a robust public access model for the dissemination of the peer-reviewed results of taxpayer (and other) funded research – the global journal corpus. Agencies dispensing funds to support taxpayer-funded research may wish to collect and publish free of charge reports generated by the recipients of those funds. However, these agencies have no rights to the research articles written for and published by journals, nor is such a claim justified by any notional absence of access. There is no evidence that making the current broad public access to the journal literature free will improve research productivity or the public wealth. On the contrary, free access, like copyright piracy, is likely to have the opposite effect.

*(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?*

Growing existing and new markets is what private sector publishers do, both for our own journals and for those we publish on behalf of our society partners.

If "market" is used in the commercial sense, publishers are growing new markets around the world as developing economies industrialize. Making publishers' and societies' content freely available is likely to stifle those opportunities, as customers choose to access free versions of journal articles rather than pay for the Version of Record. As an example, having established a growing business in China, after decades of content pirating, we now see two thirds of PubMedCentral  (PMC – the repository for mandated NIH grantees deposit) usage originating outside the US, much of it we assume from China. We also know that there are businesses in China reselling PMC content which, although made freely available from PMC, still carries a publisher's or a society's copyright. How do US taxpayers benefit from the transfer of intellectual property such as this, owned by US corporations and not-for-profit societies, to businesses and governments abroad?

There are no studies that support the notion that free access to the research literature will increase research productivity or economic growth. The fallacy is that access to the research literature by those able to make use of it is rate limiting, and that there is untapped creativity that will be released if access is made free. The modern research enterprise is complex and requires huge investments; access to the research literature is not a constraint.

We do not accept the premise that because government funds scientific research, it is entitled to full access to and control of manuscripts reporting on this research. Publishing peer-reviewed research is expensive and someone has to pay for it. The government pays only for the research; it cannot lay claim to the final publication. Having each funding agency open its database of funded projects, including research project reports and lay summaries, best serves the public interest and protects the scientific research enterprise.

Society today depends on a system of research communication that provides extremely broad access and strong quality controls. Research publishers are custodians of this system today because of the essential role that they play in the communication of scientific, technical, and medical research results. While it is the case that peer reviewers are generally not paid for performing the work of peer review, peer review is not free. Publishers invest hundreds of millions of dollars in end-to-end software tools to manage the peer review process and often also financially support the editorial groups who manage and perform peer review of submitted articles.

Government should not impose unfunded mandates that pertain to the outputs of the publishing process, including accepted author manuscripts and published journal articles. Such policies would not be justifiable or warranted. Government-imposed public access policies would violate fundamental copyright principles by allowing the government to diminish existing copyright protections for private sector journal articles.

Publishers make ongoing capital investments and incur significant operating expenses in carrying out these value-added activities.  These are not paid for by taxpayer dollars.  Any unfunded mandate has the potential to limit our ability to create the peer-reviewed literature in the first place.

*(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?*

The federal government could:

- Make funds available for the purchase of open access to published articles. Several research funders already do this (Howard Hughes Medical Institute, The Wellcome Trust, Max-Planck Institutes). These costs are a small fraction of the investment in the research itself.
- License content from publishers and learned societies and make it available to specific audiences. We license content to customers of many kinds, including government agencies, and would be pleased to enter into negotiations regarding access by specific communities to packages of content.
- Make the funder-collected and maintained outputs of taxpayer-funded research, including grant reports or research progress reports, freely available to the public. Work with private sector publishers to make that content discoverable and link it to the journal literature.

What it should not do is to take accepted or published articles from publishers or learned societies (directly or via a mandate placed on grantees) and make them freely available.

*(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?*

See the Preamble for our response to "long term stewardship".

Scholarly journal articles have been published for several hundred years by a combination of society, not-for-profit, and for-profit publishers. Together, we have provided access to the literature for scholars and researchers. A multiplicity of publishers has not prevented broad public access. In fact, one could argue that it has been an advantage in promoting competition which has, in recent years, driven development of increasingly sophisticated platforms to deliver this content.

Publishers over the past decade have developed the Digital Object Identifier (DOI) a unique identifier for each piece of content, in this case a journal article. CrossRef , a not-for-profit group founded by a group of publishers, including Wiley, in 2002, maintains 50 million DOIs. Almost 1,000 publishers and societies participate and assign DOIs to their published content items. Development of the CrossRef service has resulted in seamless navigation of the research literature by users, so that researchers using the bibliography in one article can link from a reference in the bibliography to the full text of the referenced article.

Is the government really a credible provider of these kinds of services? Given government budget constraints why would the government consider using taxpayer dollars to duplicate an existing, well-functioning service? PubMed Central, the repository for mandated NIH grantees, is not a simple archive of articles but a sophisticated publishing platform requiring millions of dollars of investment. Have the full costs of similar repositories been developed in any consideration of an expansion of the NIH mandate?

**(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?***

There are a number of projects underway or envisioned for public-private partnerships.

**Funding agency information**

Most researchers acknowledge in their publications the research funder support they have received. However, there are no standards on how authors should do this. Consequently, funders find it difficult to know and track what publications have arisen from the research they have funded. Publishers are developing a means of standardizing funder information so that this information could be made easily available to funders.  We believe that a community-wide solution of this type will be easier and far less expensive to construct than each agency developing its own response to the problem. Publishers are in the best position to provide a simple way of ensuring that journal articles are accompanied by standardized, high-quality metadata providing information about the agency, program, and even specific grant that funded the research. This proposal has been endorsed by CrossRef and major STM publishing trade associations. The technical details of implementation will be worked out through consultations with CrossRef and appropriate publisher, society, and agency representatives.  Our goal is to launch this feature in 2012.

With the successful implementation of this proposal, research funders would have access to the standard metadata from published articles that have arisen from the research they have funded. By displaying this information on their funder websites visitors will be able to follow the link (enabled

through the DOI) to the publisher's platform, where article abstracts are freely available and the Version of Record (VoR) (maintained by the publishers) is available through a variety of access mechanisms.

**DOIs for data sets**

Increasingly, investigators are being asked to share, or provide plans regarding how they will share with other researchers, the primary data and other supporting materials created or gathered in the course of their work. STM publishers and societies make significant amounts of this material available as supplementary material to published articles and are already participating in a number of initiatives designed to facilitate the sharing of data. We would be willing to work with funders and database/repository operators to develop recommended practices for assigning DOIs to data sets and supplementary material so that datasets could be linked to primary research articles.

**Author disambiguation**

Name ambiguity and attribution are persistent, critical problems embedded in scholarly research. STM publishers are working to eliminate this problem through an initiative called the Open Researcher & Contributor ID (ORCID) project (www.orcid.org). ORCID is a newly established non-profit organization whose goal is to establish an open, independent registry of researchers that is adopted as an industry-wide standard to resolve systemic name ambiguity by means of assigning unique identifiers linkable to an individual's research contributions. Researchers will be able to create, edit, and maintain an ORCID ID and profile free of charge, including defining and controlling their own privacy setting. Publishers including Wiley have contributed heavily to ORCID's initial funding.

Such a standard will not only enhance the scientific discovery process but also improve the efficiency of funding and collaboration. Participation in ORCID is open to any organization that has an interest in scholarly communications. All software developed by ORCID will be publicly released under an open-source software license approved by the Open Source Initiative (OSI). ORCID is governed by representatives from a broad cross-section of stakeholders including publishers, societies, libraries, and other institutions (see http://orcid.org/board-of-directors).

**Content mining**

Content mining has the potential to be useful to the scientific community in driving interdisciplinary research and supporting the identification of new areas of discovery. Publishers and their society partners are committed to managing content in digital formats to ensure that users gain maximum benefit. We propose to work with research funders to develop pilot projects for journal content mining that would identify, organize, and perform analysis to identify and create conceptual links within and between that content that are not obvious to initial human inspection. Although there are various ways to perform this type of processing, certain elements are common to all methods, including an automated way to process all sizes and types of content in which to identify relevant information, and facilitate its extraction and analysis.

Such pilots would focus on goals such as:
- Structuring input text, deriving patterns within the structured text, and evaluating and interpreting the output;
- Extracting semantic entities from publisher content for the purpose of recognition and classification of the relations among them; and
- Enabling developers who wish to design and implement applications to analyze our content or test applications as part of their research within publisher content.

Consensus approaches within the community could also be explored for developing better standardized, mining-friendly content formats, a shared content mining platform, and commonly agreed permission rules for content mining.

The Publishing Research Consortium recently completed a study on article-level content mining based on a broad survey of ongoing or planned activities among nearly 30 STM publishers or associations, http://www.publishingresearch.net/documents/PRCSmitJAMreport20June2011VersionofRecord.pdf

**Linking to/from research reports**

We propose a collaboration with research funders to determine whether and, if so, how publisher content could be "mapped" against research reports and other funder content. The goal would be to make connections between content items that would add value and richness to both groups' digital offerings. Specifically, this collaboration would send users from publisher websites to the funder web site to view free government-sponsored research reports, and would send users from funder sites to view free abstracts and links to the Version of Record of articles connected to a particular research report or funded project.

If successful, this will result in interoperability between funder and publisher content and would enable us to work with research funders to identify, organize, evaluate, and highlight published results from their research funding and identify relationships, projects, and offerings.

Possible outcomes of the pilot could include:
- The ability to identify all agency-funded research within publisher offerings and the ability to deliver associated metadata to that funder;
- The ability to establish mechanisms and approaches that could be implemented (for all research funders) across the industry;
- A capability to report to major funders on the impact of the research they fund, e.g. through bibliometric and other tools;
- A "research dashboard" capability or the ability to contribute to one already in existence – e.g. http://rd-dashboard.nitrd.gov/;
- A mechanism for low-cost content rental access to published articles (VoR);
- Subject area content portfolios of NSF-funded research articles for internal NSF use (e.g. study sections); and
- The opportunity to use the http://www.science.gov/ and http://www.research.gov platforms to extend this pilot to other federal funding agencies.

*(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?*

See (4) above under public-private partnerships.

**(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 awardees institutions, scientists, publishers, Federal agencies, and libraries?*

Publishers would argue that those who can benefit from access to the peer-reviewed journal literature already have access; researcher surveys bear out this assertion. We would also argue that we own and have copyrighted journal articles which are published in the journal titles we publish.

Publishers and learned societies are committed to the wide dissemination of our content. We support any and all sustainable access models that ensure the integrity and permanence of the scholarly record. This includes 'gold' open access, where publication is funded by a publication fee or article processing charge. Many publishers now offer open access options and/or publish open access journals, and work closely with funders, institutions, and governments to facilitate these developments. We believe that authors should be able to publish in the journal of their choice, where they feel their work will be best reviewed by their peers and where its publication will have the greatest potential to advance their field. Research funders could provide a fund to publishers to cover gold open access publishing fees.

**(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?***

No. Publishers also invest in these other types of content used by researchers, often by conceptualizing the project, commissioning the content, and investing heavily in its development. Any kind of mandated free access to that content is simply an expropriation of that content. The Federal Government might as well demand free access to *Time* or *Newsweek*.

**(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?***

There are no "appropriate" embargo periods. Any embargo period is a dramatic shortening of the period of copyright protection afforded all publishers.

We believe that peer-reviewed papers should not be made public within the duration of the article's copyright without the copyright holder's permission. For accepted author manuscripts and published journal articles, both of which publishers have invested in heavily, Wiley believes that publishers – and learned societies – themselves should determine the business models under which their publications operate. This should include the time, if any, at which the final peer-reviewed manuscript or final published article are made publicly available. Peer-reviewed papers are not the direct result of the Federal Government's investment. They should not be made freely available to the public unless the copyright owner authorizes the government to do so. Since the mid-1990s, the science journal publishing industry has been a key player in the dramatic digital revolution in the sciences, investing heavily to drive the shift of published research from print-only to "E-only." Rapid innovation in the publishing industry has dramatically improved functionality and efficiency for doctors and researchers, who can now perform complex searches of journals, immediately retrieve and print full text articles, link instantly to other cited articles, export text to other databases and programs, and receive e-mail alerts when new journal issues are released. Mandating free access will stifle innovation in what is now a rapidly changing environment, both by decreasing the amount that publishers are able to invest and reducing their incentive to try new approaches.