

Dr. Karen Cole, Director, and library staff
Ken Davis, Chair of Faculty Assembly Information Resources Committee, and
committee members
January 12, 2011
Archie Dykes Library of the Health Sciences, University of Kansas Medical
Center
Kansas City, Kansas

Comment 1

a) 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?

Yes.

1. Agencies can require that the results of all federally funded research be deposited into federal or institutional repositories without restrictions on access within a reasonable amount of time.
2. Agencies can provide financial support for existing and entrepreneurial efforts by publishers, individuals, and researchers in developing or improving innovative tools to mine and analyze publications in a variety of ways. Text analysis, knowledge discovery, and mapping interdisciplinary connections are market opportunities ripe for development and expansion. Open access to content (peer-reviewed article manuscripts, citations, reports, conference proceedings, and data), without subscription paywalls, is critical in order to develop such tools and reap the market rewards. While the following tools apply broadly to scientific literature, a new market could emerge that focuses on tools targeting publications resulting from federally funded research. Some examples include:
 - the work of Carl T. Bergstrom, Jevin D. West, and Martin Rosvall: <http://chronicle.com/article/Maps-of-Citations-Uncover-New/128938/>
 - Maps of Science, developed by SciTech Strategies, Inc.: <http://mapofscience.com/index.html>;
 - the work of Katy Borner of Indiana University: <http://ella.slis.indiana.edu/~katy/>
3. Agencies can support and expand access to peer-reviewed publication data by maintaining their own repositories and providing metadata. For example, PubMed now has over 21 million citations and, particularly since NIH's open access requirement of 2007, provides increasing full-text access to taxpayer-funded research. Citation databases such as PubMed provide enormous datasets but without open access to full text content, potential markets and the entrepreneurial need for new tools will not be fostered. An example of an existing tool that might benefit from additional expansion and financial support is [Science.gov](http://www.science.gov) -

<http://www.science.gov/>. Many of the participating agencies in this endeavor could use this tool to make available the final research reports resulting from research the agencies have funded.

b) 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?

One example of a policy that has already shown potential to grow the economy and improve the productivity of the scientific enterprise is the NIH Public Access Policy (<http://publicaccess.nih.gov/>). In 2005, the NIH offered a voluntary policy requesting that scientists deposit into PubMed Central a digital copy of any research findings published in a peer-reviewed journal that were obtained using NIH funding. Compliance rates were less than 4%. Since the NIH Public Access Policy went into effect in 2008, the rate of article submissions has jumped to 77%. This demonstrates that federal mandates requiring free and open access to the results of federally funded research can greatly increase the amount of research content available to the public.

The key to policy mandates is that published results of federally funded research must be openly available in a digital archive so that results can be archived and reused. It is the ability to preserve the content and to reuse that content in derivative ways that will allow others to make innovative uses of federally funded research findings.

Agencies should take bold steps in ensuring the timely reporting and review of research that they fund. A recent study by Yale School of Medicine found that fewer than half of a sample of clinical trials funded by NIH were published within 30 months of completing the clinical trial (Ross et al. Brit. Med. J. doi: 10.1136/bmj.d7292). Alternatives to journal publication exist for cases where journal publication is either slow, problematic, or not ideal. These alternatives (e.g. ClinicalTrials.gov) are already widely supported by agencies.

References:

Many NIH-funded clinical trials go unpublished over 2 years after completion. http://www.eurekalert.org/pub_releases/2012-01/yu-mnc010312.php

UNIVERSITY RESEARCH: Most Federal Agencies Need to Better Protect against Financial Conflicts of Interest. 2003. p. 8-12. <http://www.gao.gov/new.items/d0431.pdf>

c) What are the relative costs and benefits of such policies?

Relative costs of the NIH Public Access Policy, as noted by Dr. David J. Lipman, Director of NCBI:

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

- Start up costs estimated at \$500,000
- It costs \$3.5 - \$4.6/million annually (on a \$30 billion budget) to provide access to results of all of their funded research.
- An investment of about 1/100th of 1 percent of NIH's overall budget results in access to 2.2 million articles.
- PubMed Central is currently used by more than 500,000 users per day, with the majority of users coming from domains outside of education – underscoring deep demand for this information throughout the public sector.

The cost to utilize an existing tool, [Science.gov](http://www.science.gov), to disseminate research results for FY 2011 was estimated at “\$67,000 plus in-kind contributions of agencies.” Source:<http://goo.gl/WD7A5>

Other federal agencies that award research dollars can use the NIH as a model for costs and benefits and see that the costs are minimal compared to their overall budget and the numbers of users show that the publicly available content is in high demand.

A policy that requires the deposit of peer-reviewed manuscripts into a single open access repository, either centralized or institutional, will achieve maximum research impact for authors and provide the broadest means of access

The costs and benefits of providing open access have been extensively analyzed, country by country, in the following reports:

Houghton, J.W. & Oppenheim, C. (2009) The Economic Implications of Alternative Publishing Models. *Prometheus* 26(1): 41-54

Houghton, J.W., Rasmussen, B., Sheehan, P.J., Oppenheim, C., Morris, A., Creaser, C., Greenwood, H., Summers, M. and Gourlay, A. (2009). *Economic Implications of Alternative Scholarly Publishing Models: Exploring the Costs and Benefits*, London and Bristol: The Joint Information Systems Committee (JISC)

Houghton, J.W. and Sheehan, P. (2009) Estimating the potential impacts of open access to research findings, *Economic Analysis and Policy*, vol. 39, no. 1, pp. 127-142.

d) What type of access to these publications is required to maximize U.S.

economic growth and improve the productivity of the American scientific enterprise?

The ability to freely access, read, distribute, re-use and modify federally funded research electronically is required in order to maximize the benefit to patient care - starting in the research laboratory and culminating at patients' bedsides through disease discovery, diagnosis, treatment and cures. Federally funded research content must be fully and freely open in order to reap the benefits of publication models that encourage publicly funded research findings to be reused. Full access should be defined as making content available outside of firewalls or other barriers that restrict text and data mining. Publicly funded research should be made available for further uses through sharing and integration in ways that take advantage of computer-readable language markers. It is the ability to fully reuse publicly funded content that can stimulate the developments in new markets.

Only providing open access to peer-reviewed content will motivate powerful new developments in cross-disciplinary research and knowledge discovery. There is only limited incentive for developing powerful new tools now in both public and private arenas, while the content is still so sparse.

Comment 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?

In accordance with U.S. Copyright Law, authors automatically retain copyright unless they transfer it to another entity, such as a publisher. In order to keep key rights to publicly funded research, one option for authors is the use of a Creative Commons license in order to select a range of terms for how others can use their content. With a creative commons license, authors select the terms of use. Terms of use can be selected that require all modifications and reuse attribute the author as the original source. The author will retain copyright while granting Creative Commons terms of use that can encourage reuse and derivative uses that can potentially grow into new products and services that are based on the findings of taxpayer funded research results.

Another option is to use the Scholar's Copyright Addendum Engine developed by Science Commons and SPARC found at this link <http://scholars.sciencecommons.org/>

Using this addendum, the author will sign an agreement that modifies

any publisher copyright language that restricts article use and reuse in ways that prohibits free public access to taxpayer funded research. The author can attach this addendum to the publisher copyright transfer agreement and sign the publisher agreement only after the publisher agrees to the terms in the addendum. The main point is for the author to retain all rights to distribute, modify, deposit, and reuse the peer reviewed, edited manuscript as accepted for publication. Publishers will be able to retain copyright of the manuscript as it appears in their publications but not the author manuscript after edits and peer- review.

Embargo on access, another protection mechanism, might be applied sparingly to protect intellectual property interests.

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?

There are publisher copyright transfer agreements which state that the publisher owns the article and controls the systems in which an article can be deposited and accessed as well as controlling how an article can be reused and distributed. Public access policies should not adopt any language that grants publishers all rights to federally funded research that results in peer reviewed scholarly journal articles as they appear now or in any unforeseen format into the future in digital systems operating now or in any systems not yet developed. Such language is currently used in several copyright transfer agreements by scientific publishers and sets extreme restrictions on how taxpayer funded articles can be accessed, re-used, distributed and archived. Tax payer funded research results should be openly and freely available in order to build upon the knowledge and results in order to offer new and innovative services that serve the public good.

Authors should retain full copyright for, ownership of, and license to a peer-reviewed final manuscript. The publisher's final version should be considered the publisher's version of record with full license for the publisher to sell, license, and build additional services around. No policy should be made that restricts either the author's license or a publisher's license to do what they want with their final versions.

Comment 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?

Subject repositories such as PubMed Central and arXiv, and institutional repositories, such as The University of Kansas Medical Center's Archie, and Scholar Works at the University of Kansas, are designed to

- enable free and open access in order to encourage and promote innovation
- benefit the public through shared knowledge, and
- encourage commercial and educational progress. arXiv is an example of a Physics subject specialty repository

We share a vision of broad dissemination of rapidly released scientific research available freely to all.

Real-world experience has shown that centralized approaches (e.g. PubMed Central) to public access can provide stable, persistent access, some threshold of quality metadata, and curation. However, the same centralized approaches lack agility and responsiveness when it comes to providing new and improved services around content.

The web is by nature decentralized. It is “distributed content provision and central harvesting, Google-style. It is not, as in paper days, that all the content needs to go in one central physical space.”

(Harnad, <http://openaccess.eprints.org/index.php?/archives/341-guid.html>)

Innovation and commercial opportunities will arise from smaller communities or commercial ventures in the interest of solving specific needs. Decentralized approaches, public and private, tend to provide innovative and more specialized services much faster. Moreover, institutions have vested interests in facilitating and ensuring deposit of peer-reviewed manuscripts into institutional repositories.

Institutions that do not manage their own repository systems still have multiple options for creating repositories. Institutions can partner with other institutions, consortia or organizations such as Dryad (<http://datadryad.org/>). Alternatively, they can license commercial repository services such as BioMedCentral's Open Repository service <http://www.openrepository.com/> or Berkeley Electronic Press' DigitalCommons <http://digitalcommons.bepress.com>.

Centralized repositories or even commercial entities, having harvested peer-reviewed manuscripts, reports, and other content from institutional repositories could provide the following enrichment services: quality metadata, canonical access points through resolvable persistent identifiers (e.g. DOI, PURL, or PMCID), support of author and institutional identifiers (e.g. ORCID), enhanced or cross-disciplinary discovery services, and distribution channels through standard protocols

such as OAI-ORE, Atompub, and SWORD.

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?

Federal agency custody can ensure constituents and stable access and use conditions that permit all interested individuals, organizations, institutions, corporations, and entrepreneurs to build upon, reuse, create new services and products, share with and educate others based upon federally funded research findings.

Although multiple private sources may certainly ensure long-term preservation they by nature will have a financial interest in controlling access. This cannot be equated with public access. The two are at odds.

Comment 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?

Currently, there are few models because of the lack of open access to publisher archives. And, while archives are immensely valuable, we must first and foremost open up access to current research. Advancements in biomedicine, physics, chemistry, and related disciplines occur too rapidly.

One model is NLM's Back Issue Digitization Project using PubMed Central - a partnership between the National Library of Medicine and publishers. This project includes archival content that was previously only available in print, in addition to more recent material provided voluntarily by participating publishers. Content is now available digitally and free to the public. <http://www.ncbi.nlm.nih.gov/pmc/>.

Another is BioMed Central where value-added services and content curation are vital to the BioMed Central private enterprise, yet journal content is freely available to all because research or institutional funds have paid manuscript submission fees.

Comment 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?

Interoperable search, discovery, and analysis can only work if content is accessible. Further, the current situation works against libraries and institutions who must pay not only for subscriptions to journals but also for additional privileged search and discovery services, many now acquired by publishers, that can access article text. Only a handful of institutions can afford this.

Only providing open access to peer-reviewed content will motivate powerful new developments in interoperable search, discovery and analysis tools across repositories and disciplines. There is only limited incentive for developing powerful new tools now in both public and private arenas, while the content is still so sparse.

1. All agencies must mandate that all federally funded research be made openly accessible without restriction except in rare cases where an embargo period is absolutely necessary.
2. Agencies must require that research findings, reports, and peer-reviewed manuscripts be deposited into an openly accessible repository.
3. Federal agencies, publishers, and academic societies should co-develop and reuse existing standards for the application of metadata and the encoding of both metadata and works.
4. Metadata is not enough. Agencies can ensure access to full-text. Institutions, commercial ventures, publishers, and societies can provide access to the XML representations of peer-reviewed manuscripts. Alternatively, publishers and other commercial ventures could develop additional editing and creation tools to assist authors and institutions in creating publications that begin life as standardized semantic documents, e.g. the NLM Journal DTD is a candidate for encoding journal content. These documents would be ripe for analysis and mining.
5. Providers should avoid designing interchange standards and instead implement common modern ReSTful, Linked Data, and Semantic Web architectures. They should start by providing small prototype data services, releasing enhancements early and often, and be open and attentive to the needs and wants of real-world data consumers and application developers.

What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities?

Open access to full-text content is vital for truly accelerating scientific advancement, dissemination of knowledge, and new market opportunities. XML representations of full-text articles would be ideal, facilitating the application of semantic web technologies to analyze and mine publications.

The OAI core metadata – author, date, title, publication, etc. – are the minimum metadata required for interoperability. They can be enhanced and made more powerful; the urgent priority, however, is not to enrich the metadata but to provide the OA content itself. OAI is more than enough for most uses of peer-reviewed journal articles – by researchers, harvesters, and the public. What is needed is the articles themselves. And for that, deposit must be mandated. Once the OA content is there, the hard part is done: Further enriching the metadata and capabilities is the easy part and will be taken up by many skilled and creative developers and metadata specialists.

Enhancements to minimal metadata should include unambiguous unique identifiers (ORCID, I-2, publisher identifier, publication identifier), copyright status, reuse rights, publication status, work type, archived-version-type, archive status. Machine-readable metadata values should be expressed as endpoints, dereferenceable identifiers or URIs.

All players, agencies, publishers, societies, institutions, and authors have an interest in measuring use, readership, and citation of works. They should work together to adopt a common standard or mechanism for measuring reuse across access points, whether a work is downloaded from an institutional repository or a publisher's site. This could mean ensuring that a registered identifier similar to a DOI is applied consistently across versions of works and is traceable across the Internet. This would provide further opportunities for businesses offering analytics services.

Comment 6

How can Federal agencies that fund science maximize the benefit of public access policies to U.S. taxpayers, and their investment in the peer-reviewed literature, while minimizing burden and costs for stakeholders, including awardee institutions, scientists, publishers, Federal agencies, and libraries?

Federal agencies can maximize the benefit of public access policies by ensuring that all stakeholders mentioned do not have to pay multiple times for access to the same peer-reviewed literature. Under the current system, taxpayers are funding research – but then have to pay again to access the scholarly publications that result from the research. Libraries license journal databases to access the work of their own researchers. As authors are increasingly being asked to pay author fees to have their work published in journals that provide immediate, free access to anyone worldwide, it is crucial that the open publishing model is truly open. Federal agencies can adopt public access policies that promote the ability to reuse and produce derivative works of the peer-reviewed literature, as well as archiving and other types of

machine readable analysis and preservation of those works.

Interdisciplinary collaboration is a critical factor in many researchers' ability to win funding. This work cannot happen without fast dissemination of new research, especially to those outside of our universities. All parties involved in collaboration need to have access to existing research findings that resulted from federal funding in order to grow and build upon those findings.

Public access policies can minimize the burden and costs to stakeholders by moving away from an ownership model where authors are required to sign over their copyrights to publishers, to a model where full open access and markets can co-exist (perhaps in the form of licensing access, with authors retaining copyright over their scholarship). Both researchers and publishers can continue to thrive. Publishers can offer many value-added services to re-package the results of federally funded research (both the peer-reviewed literature and accompanying data), for which there would be a strong market. Additionally, if authors retain ownership and grant terms of use that encourages reuse and derivative works rather than restricting such activity, new markets and existing publishers could potentially open to them and offer them value-added services that currently do not exist. Libraries would likely have more resources available to purchase those value-added services if they were no longer locked into restrictive journal license packages with prices that vastly exceed the rate of inflation and budget increases.

A second important step is to ensure that any policies adopted are consistent across all Federal agencies. Awareness of the policies must be raised using clear, simple language. Policies must be easy for all stakeholders to understand and follow. Policies should include a uniform set of instructions and tools must be developed to facilitate the compliance process.

Comment 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?

Yes, all peer-reviewed publications, regardless of format, resulting from federally funded research should be covered by public access policies that make them openly accessible without restriction to the public. However, we recognize that different publication types may require different policies to account for copyright, royalties to authors, format and access issues. Books, for example, generate author royalty fees while scientific journal articles do not. Where possible, if the author

and publisher both agree, the deposit of publications – over and above refereed journal articles and refereed conference papers – such as book chapters, books and research data should be encouraged as well. It is important that these other types of peer-reviewed publications be deposited in an openly accessible repository, with standards for metadata developed and applied, so this valuable content can be more easily located and used.

We recommend that priority be given first to making scholarly journal articles resulting from federally funded research freely available. That is something that most authors want, whereas there is not wide agreement that book content should be made open access.

Comment 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?

Ideally, immediate, free open access is desirable so that the public can make use of the research and start building upon it and making innovative uses of it that serve the public. Embargo should be the exception. It should be applied only when a strong case can be made that it is needed to protect the author, a pending patent, the funder, or the public good.

Please describe the empirical basis for the recommended embargo period.

The sooner that research is released, the sooner the public can benefit from the findings and economic growth in new markets and services can be developed. Since the NIH mandate took effect in 2008, publishers have accepted the 12-month embargo and it has not been proven that releasing content freely after twelve months results in publisher loss of revenue. Because statistics show that the immediate release of published research results in higher citation counts and more widespread use of the content, an embargo of less than twelve months or a zero embargo would have a higher impact on the ability of the public to use the content that results from publicly funded research.

Other items the Task Force might consider for Federal policies related to public access to peer-reviewed scholarly publications resulting from federally supported research:

1. Consider using Federal policies to expand upon what appears to be standard practice for several government agencies - providing public access to unclassified research reports at the conclusion of the funding period. See David Wojick's arguments for this

at<http://scholarlykitchen.sspnet.org/2012/01/06/my-argument-for-public-access-to-research-reports/> . Examples of currently supported information portals and content include:

- a. Department of Energy: <http://www.osti.gov/bridge/>
 - b. National Science Foundation: <http://www.research.gov/>
2. Enforce existing mandates to “ensure that all final project reports and citations of published research documents resulting from research funded, in whole or in part, by the Foundation, are made available to the public in a timely manner and in electronic form through the Foundation's Web site.” Source: America COMPETES Act of 2007, H.R. 2272, 110th Cong., 2007
- a. Sec. 7010. Reporting of Research Results.http://www.nist.gov/mep/upload/PL110-69_8907.pdf

Karen Cole

Director of Dykes Library
University of Kansas Medical Center