To: Office of Science and Technology Policy (digitaldata@ostp.gov)
From: Public Library of Science (PLoS)
Email: dokubo@plos.org
Subject: Response to the OSTP RFI on Public Access to Digital Data

Date: January 11, 2012

The Public Library of Science (PLoS) welcomes this opportunity to respond to the Office of Science and Technology Policy's Request for Information (RFI) on the topic of Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research. PLoS is a nonprofit publisher and advocacy organization with a mission to accelerate progress in science and medicine by leading a transformation in research communication.

PLoS was founded in 2000 by three scientists: Harold Varmus, President of Memorial Sloan-Kettering Cancer Center in New York; Patrick Brown, Professor, Stanford University School of Medicine and Howard Hughes Medical Institute in California; and Michael Eisen, Associate Professor at the University of California, Berkeley. PLoS became a publisher in 2003 and currently publishes seven peer-reviewed fully open-access journals. They are *PLoS ONE*, which publishes all rigorous science across the full range of life and health sciences; the "community" journals (*PLoS Genetics*, *PLoS Computational Biology*, *PLoS Pathogens*, and *PLoS Neglected Tropical Diseases*); and our flagship journals, *PLoS Medicine* and *PLoS Biology*, highly selective journals publishing fewer than 10% of submissions along with a range of informative and influential non-research content. From publishing only 15–20 articles a month when *PLoS Biology* was launched in 2003, the PLoS corpus now comprises almost 42,000 peer-reviewed articles, most of which are the research output of public funding. This is accomplished by 115 professional staff with the expert assistance of more than 2,500 academics on the journal editorial boards and tens of thousands of reviewers. The huge support we have garnered from the academic community is a testament to the success of PLoS as a publisher and to full open access as both a business model and a model that is changing the landscape of scholarly communication. While our core business is publishing, our key objectives are to:

- Provide ways to overcome unnecessary barriers to immediate availability, access, and use of research
- Pursue a publishing strategy that optimizes the openness, quality, and integrity of the publication process
- Develop innovative approaches to the assessment, organization, and reuse of ideas and data

We are pleased to provide the following specific comments in response to your questions.

(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?

Comment 1

1.a Growing existing and new markets related to access.

The greatest impact that agencies can have to grow new and existing markets is to mandate that their grantees make the published output of their research immediately and publicly available for anyone to reuse using an appropriate open-access license. Open access enables third parties to not only read the literature online for free ("free access"), but also to reuse that content without restriction—including commercial restriction (hereafter termed "full open access"). Only about 8% of the scientific literature is currently available under an open-access license [1]. This means that up to 92% of the scientific literature is not available as a market opportunity to build new and innovative tools that will 1) increase the speed of output and productivity of scientists, 2) start new businesses and foster future startups, and 3) generate more jobs.

It is important to understand the implications of the distinction between free access and full open access. A license that permits others to read the article for free but that does not allow derivative reuse, including commercial reuse, is inadequate and will not realize the full potential of the US scientific enterprise. The license applied to the published work must not, therefore, limit the legal right for third parties to make derivative copies or reuse it in any way (as long as appropriate attribution is given). It is also important to be aware that third parties consist of computers as well as human readers. Computers function more effectively when browsing text on the open web [2], e.g., for text- and data-mining, and derivative reuse, such as translation to another language, as well as the open exchange of metadata and application programming interfaces (APIs). This will enable markets beyond that of basic science and publishing to benefit from the output of agency funding. A license that enables reuse in any form, including commercial reuse, is the Creative Commons Attribution License (CC-BY, [3]), used by PLoS and other open-access publishers (e.g., BioMedCentral).

As Kim Hailey, Founder of Transmute, Inc., PixelCam, Inc., and Meta-Software, Inc., put it (personal communication):

> *As a founder of an advanced energy start up, we depend on open access. Innovation relies on unfettered access to a broad range of subjects, giving rise to groundbreaking, multidisciplinary discoveries. It is important that our researchers have access to the latest articles as this ultimately shortens the time to market for new products. Information should not be hidden behind publishing company pay-walls. Even worse are the articles published in conference proceedings that are only available to those who attended the conference. Cost is only part of the issue. The problem is complicated by having to maintain many journal access accounts and to duplicate them for all members not at the same site as is the case in most virtual organizations. Much of the research is government funded and should be electronically accessible to all. The best way of leveraging the public investment in scientific research is to demand that it be openly accessible to all*.

Funding agencies should therefore encourage academic scholarly publishers to see themselves as service providers, rather than product providers where the product being sold is access to the output of research. Making the published work open access is a service for research funders, scientists, patient advocacy groups, educators, and other businesses (e.g., pharma and agriculture). Open access incurs a

cost and should be paid for, but once the publication fee is paid, publishers do not need to hold the rights to, or demand further payment for, access.

1.b Growth of new tools to analyze peer-reviewed publications—article-level metrics and innovation.

Full open access to the literature (and the associated data) can help promote the development of appropriate analytical tools, such as article-level metrics (ALMs), to enable scientists and businesses to target and track more accurately the research findings they need to grow. ALMs also provide a means for federal funding agencies to have a range of metrics to evaluate their output. The development of these metrics provides a compelling example of the potential power of open access to stimulate the growth of innovative tools that can help speed the progress of science itself and facilitate its application by third parties. Such innovation has the potential to directly foster job creation, but only if a sufficiently large corpus of papers are available as open access—the 8% of the market share currently available as open access is insufficient. Federal agencies can take a lead by supporting the growth of open access and the development of tools such as ALMs.

ALMs are an example of a new category of tools to help analyze peer-reviewed publications. Researchers, funders, and readers are interested in metrics at the article level so that they can identify and curate the most relevant research publications [4]. Traditionally, only one metric, the journal impact factor, has been used as an indicator of the merits of a research publication, but a range of metrics are likely to be more informative [5,6]. By studying the citations to the individual articles, for example, a scientist or a funding agency can understand in a more nuanced way the research impact of that article. The journal *PLoS ONE,* for example*,* has some papers that have received more than 200 citations, while 20% of articles one year or older have received more than nine citations and 76% one or more, a pattern of variation that will be familiar to all journal publishers. Impact also takes many different forms, so qualitative as well as quantitative metrics need to be developed. An article published in 2008 by *PLoS Medicine*, for example, provided a "Dirty War Index" [7] that has been adapted for use in NATO military environments, such as southern Afghanistan, to reduce the possibility of injuring Afghan civilians. The approach has led to NATO changing procedures [8]. Citations or the number of downloads cannot reflect this impact. Scientists or individuals from industry may also want to have the means to track research threads using social networking tools—who is reading what, where, and when, and what are those scientists then publishing or patenting. Such data-analytic information is useful for markets such as biotech and agriculture, as well as basic science—each will want to be able to identify the most relevant paper for their business or research.

Of course, innovative tools for research evaluation needn't just come from publishers—they can also be developed by third parties, thus creating new markets. But the effectiveness of such tools requires that published papers have a fully open-access license that legally allows reuse by others (such as CC-BY) and the technology to track that reuse. Not having more than 90% of the papers available as open access is therefore vastly limiting this market potential. Several software applications produced by individuals in response to a public competition hosted by PLoS and a private company called Mendeley [9], for example, created prototypes to measure the impact of articles or researchers (see also Comments 1.c and 4 below). The effectiveness of such tools is hampered because they apply to such a small share of the literature.

The same CC-BY license should apply to the ALMs themselves. For example, if ALMs were to gain acceptance, then it could be conceivable that a not-for-profit entity could take over the project or a new

proprietary company be created that specifically focuses on the production of ALMs. This would potentially hinder the speed at which these evaluation and tracking tools develop.

1.c Growing the economy and improving the productivity of the scientific enterprise.

Full open access promotes scientific productivity because there is no delay in accessing the information from a research publication and it does not impose any restriction—even commercial—on any derivative reuse. This facilitates text- and data-mining, or the extraction of tables, figures, and supplementary materials from the published paper. It also permits the published work to be translated into a different language or for that work to be included in marketing information for biotech, pharma, and other industries, as well as in textbooks or course packs for educators or patient advocacy groups.

There is increasing evidence that open access increases the rate at which articles are cited [10,11]. Scientists can read and extract up-to-date information, but so too can small biotech companies, small agricultural businesses, fisheries, and public health and patient advocacy groups or charities. This promotes a level playing field among those seeking information and engenders competition. Regardless of the subject you are interested in, the size and wealth of the institute or industry you belong to, there is no barrier to your obtaining and analyzing the information you need:

> *This flood of data and analytical opportunities creates more value for people who can be creative in seeing patterns and for people who can be entrepreneurial in creating new business opportunities that take advantage of these patterns. My hope is that the technology will create a platform that people can tap into to create new entrepreneurial ventures—some of them, perhaps, huge hits like Facebook or Zynga or Google. But also, perhaps equally important for the economy, hundreds of thousands or millions of small entrepreneurial ventures, eBay based or app based, would mean millions of ordinary people can be creative in using technology and their entrepreneurial energies to create value. That would be an economy where not only does the pie get bigger but each part of the pie—each of the individuals—benefits as well*. (Quote from MIT Professor Erik Brynjolfsson [12].)

An example of such third party creativity bringing together analytics and scientific innovation is the outcome of a partnership between PLoS and Mendeley [9] to host a public competition—the PLoS/Mendeley Binary Battle—to build software applications (apps) that make science more open and useful for the reader. Participants were asked to make use of PLoS and Mendeley content and the freely available API provided by PLoS and/or Mendeley [13]. Mendeley is a company that enables individuals to organize their reference library and PDFs online and to find and share information with others that have similar interests. The judging panel consisted of Tim O'Reilly, founder of O'Reilly Media, Juan Enriquez, Managing Director of Excel Venture Management, John Wilbanks, former VP for Science at Creative Commons, James Powell, CTO of Thompson Reuters, and Werner Vogels, CTO of Amazon.com. The top two winning applications showed prototypes of tool applications for productivity, analysis, and decision tools for faster and better scientific communication. For example, openSNP allows customers of direct-to-customer genetic tests to publish their test results, find others with similar genetic variations, learn more about their results, find the latest primary literature on their variations, and help scientists to find new associations [14]. PaperCritic offers researchers a way of obtaining and providing feedback for each other's work in a fully open and transparent environment [15]. Other, shortlisted applications provided a range of qualitative and quantitative tools to measure the impact of articles or researchers (e.g., [16–18], see also Comment 1.b).

Social media, video, on-site aggregation, and metric tools can all be used to discover, bridge, and respond to new unforeseen markets if research content is fully open access, such as when there is an urgency for the rapid exchange of data and information. To facilitate this, PLoS (in partnership with the National Center for Biotechnology Information [NCBI] and Google Knol) has developed a new publishing platform—PLoS Currents—specifically for especially rapid dissemination. PLoS Currents: Influenza, for example, was launched during the 2009 flu pandemic [19,20]. A more recent launch, PLoS Currents: Disasters, caters to the study or management of all types of disasters, particularly for data and analyses that might not otherwise be openly shared [21].

When research publications are not silo-ed into subject-specific journals behind access barriers, they can be collated from multiple sources, regardless of discipline or the journal in which they are published, and then made available on a different platform. With appropriate technology, such "hubs" of information can be developed to cater to different communities and markets and be enhanced via communication and data-analysis tools (such as that described in the PLoS/Mendeley Binary Battle). The simplest way to organize content is to package relevant articles into open-access subject-specific collections. PLoS has a range of collections [22] covering papers from all their journals, some of which are the outcome of specific collaborations (such as with the Census of Marine Life, a ten-year initiative intended to assess and explain the diversity, distribution, and abundance of marine life in the world's oceans [23]; see also Comment 4). Another solution is to provide hubs of activity around certain topics. One such initiative, still in the early stages of development, is the PLoS Biodiversity Hub [24], funded by the Sloan Foundation, which allows individuals from the academic community ("curators") to select and filter articles, regardless of the journal of origin, and which can be enhanced by comments from curators and via semantic linking. Again, aggregation, curation, and annotation are only possible because the articles are fully open access. The fact that about 92% of the literature is not fully open access (see Comment 1a) again severely limits the growth of this market.

Open access, therefore, improves the productivity of the scientific enterprise by making information immediately available and without reuse restrictions. With full open access, PLoS and other new and existing companies have already created, and will continue to create, innovative new initiatives to increase productivity of scientists and businesses.

1.d The costs and benefits of open-access policies.

According to the Houghton Reports [25–27], open access is likely to return a 5-fold increase in investment. Their analysis shows that the benefits of an open-access policy greatly outweigh the costs (e.g., the benefits of the National Institutes of Health [NIH] policy are estimated to be about eight times larger than the costs). Expanding an NIH-style policy to all other US science agencies is expected to bring in an additional $1.5 billion in revenue, of which 60% is estimated to benefit the US economy directly.

Although much of the additional revenue will be realized via innovation, increased scientific productivity, and the growth of new markets, it will also promote greater competition among publishers to provide a more cost-efficient service for authors and funders [28,29]:

> *The proliferation of online OA journals in combination with aggressive consortia licensing would significantly alter the current business model of academic journal publishing. The creation of OA electronic journals is a form of entry into the academic publishing industry. By multiplying the number of journals available not under the control of for-profit publishers, OA publishing would*

> *increase competition within the industry as well as increase the bargaining power of academic libraries and faculty authors. As the use of e-journals becomes more accepted, traditional publishers would most likely be forced to change their role. Rather than acting as oligopolists that profit by controlling access to a small number of prestigious journals, they may be forced to act as agents of the libraries, negotiating with journal providers and packaging e-journals as requested by the libraries*. [29]

Archiving and maintaining access to previously published literature will incur a cost. PubMedCentral (PMC) is a repository of freely available published articles maintained by the NIH and used by more than 500,000 users per day, most of whom are not involved in education. NIH reports that it costs $3.5–$4.6/million annually (on a $30 billion budget) to provide access to the results of all of their funded research [26]. This means that an investment of about 1/100th of 1 percent of NIH's overall budget results in access to 2.2 million articles.

Other federal agencies can use the resources and infrastructure (such as PMC) already developed by NIH to implement a consistent policy on access across all the federal agencies. Avoiding duplicating effort will help to minimize costs.

1.e A full open-access license is the only one to endorse.

As we have discussed extensively above, the only license that can really liberate the full market and scientific potential is one that permits reuse in any form, including commercial reuse. This is full open access, as provided by the Creative Commons Attribution License (CC-BY) or similar. This license guarantees the creator of the work the legal right to appropriate attribution, while allowing for unrestricted reuse by the public.

The license applied to the published work sets out the conditions under which that work can be distributed and used. The license is distinct from copyright. The copyright holder determines the terms of the license and holds the rights to the article under those terms. Ideally, the copyright holder should also be the author of the work. Copyright holders such as publishers or funders should not, however, restrict reuse, but allow users to "copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship".

This is explained in the following extract of an article published in *PLoS Biology* [30] by Professor Michael Carroll *(*Professor of Law, Washington College, Director, Program on Information Justice and Intellectual Property and Member of the Creative Commons Board):

> *Granting readers full reuse rights unleashes the full range of human creativity to translate, combine, analyze, adapt, and preserve the scientific record, whereas traditional copyright arrangements in scientific publishing increasingly are inhibiting scholarly communication. Traditional copyright law was designed with the subscription-based publishing model in mind. Authors receive copyright when they write their first draft of an article. Authors then transfer this copyright, or grant an exclusive license, to a publisher in exchange for publication. The publisher relies on copyright to police the behavior of readers and competitors who may seek to obtain or redistribute the content without a subscription.*

By shifting the financing away from subscriptions, the open-access model realigns copyright to enable broad reuse while assuring authors and publishers that they receive credit for the work they have done. This is done through open licensing by the copyright owner. Initially, the authors of an article automatically own a copyright in the article as soon as it has been drafted. If the authors sign an agreement that transfers the exclusive rights to the publisher, the publisher becomes the copyright owner. The standard means for achieving open access with respect to copyright is for the copyright owner (author or publisher) to use the Creative Commons Attribution license [3], which gives readers and republishers broad reuse rights on the condition that credit for the article is given as directed by whoever is granting the permission. (Disclosure: I sit on the Board of Creative Commons.)

Recently, however, some commercial publishers have waded into the open access waters by charging authors a publication fee to substitute for subscription revenue while limiting reuse. Having been paid for coordinating peer review, editing and laying out the text, and the like, these publishers nonetheless limit readers to making only non-commercial reuses, or even also requiring reusers to use the same license for any adaptations, while reserving to the publisher the rights to make any commercial reuse. (This is done through use of the Creative Commons Attribution Non-Commercial license or the Creative Commons Attribution Non-Commercial Share-Alike license.) This is pseudo open access. Authors who pay for publication in these pseudo open access publications are not getting their money's worth. For example, text or figures subject to these more restrictive licenses cannot be uploaded to Wikipedia, which uses the Creative Commons Attribution Share-Alike license.

Presumably, these publishers retain commercial reuse rights either to derive additional revenues from certain potential reusers or to block competitors, who may exercise these reuse rights to earn revenue through some kind of value-added service or publication. This latter option is possible only if the competitor discovers a market that the original publisher overlooked. Such entrepreneurs should be rewarded rather than controlled.

Federal funding agencies should insist on a full open-access license, such as the Creative Commons Attribution License, to ensure that there are no commercial barriers to future entrepreneurs.

(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?

Comment 2

As noted at the end of Comment 1.a, publishers should see themselves as service providers, overseeing the production and peer review of articles and making them available under a full open-access license. This is a service that incurs a cost and for which they should be paid. They should not, however, be allowed to retain rights to the future use of that article once the service has been provided.

An additional point to consider is that editors and reviewers of published articles are mostly federally funded academics who are not paid for their services by publishers. Yet they are also stakeholders in the publication process, as they have made an intellectual investment in safeguarding the scientific rigor of the research article. If publishers insist on retaining access rights based on intellectual property arguments, which we do not agree with, then the right of these other service providers also needs to be taken into account, especially given that it is primarily federal funding agencies that pay for their time.

Federal funding agencies should therefore ensure a license is applied to their publications (and associated metadata) that enables unlimited reuse. As discussed in 1.d and 1.e, the Creative Commons Attribution License (CC-BY), or similar, ensures the legal right for author to be attributed for the work while permitting any derivative reuse. By providing legal protection for appropriate attribution, the intellectual property interests of scientists and federal agencies are appropriately protected.

(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?

Comment 3

The research output of federally funded research should be preserved and useable in perpetuity and held in publicly available repositories. Such research should never be held as proprietary data with limited access and reuse for all the reasons mentioned above, so it is vital that publishers are not solely responsible for archiving the scientific output of federally funded research. All repositories should also apply an appropriate license to their content to ensure public accessibility and reuse by third parties (e.g., CC-BY).

Central repositories can provide advantages for both federal government and third party users. For example, federal agencies can take advantage of other tools and services they provide. In the case of NIH, PMC can provide links to other data sources held by NCBI. OpenAIRE [31], a European central repository across a network of 27 countries, is currently being set up to help researchers comply with European Community policies on open-access deposition and will also provide additional services. The aim of this initiative is both to ensure the long-term preservation of the literature, regardless of which member state the article originated from, and to develop the technology and infrastructure to enable the data associated with publications to be manipulated. Central repositories are also likely to prove cost-effective, as they minimize duplication of effort (see Comment 1.d).

Full open access also facilitates distributed archiving across a range of independently held repositories. Third parties can therefore set up and maintain independent repositories, promoting public–private partnerships. For example, particular communities or businesses may want to aggregate and archive specific content and enhance this with subject-specific tools and analytical services (similar to the rationale for PLoS Hubs). The federal government should ensure that such third party repositories also

meet conditions for reuse and public accessibility (e.g., the Wellcome-funded United Kingdom mirror of the NIH-funded repository PMC). Repositories that hold the content but do not permit reuse are not a viable option, as it is only through reuse that effective long-term preservation can be maintained.

Both central and distributed repositories are likely to enhance and ensure the long-term preservation of publications. Ensuring the content is open access permits both types of repositories to be developed in parallel. It is important to maintain interoperability and data standards between repositories to ensure federated searching and discovery across all content. Consistent licensing that permits reuse of the content will also provide incentives to create and preserve interoperability.

To ensure the preservation of the scientific record, the federal government should also, at a minimum, maintain a publicly accessible mirrored version of all the content and ensure that the content is useable over the long term (including machine reading) despite changing technology.

(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?

Comment 4

4.a Models for public–private partnerships.

With its mission of leading a transformation in research communication, PLoS is well-positioned to provide a test-bed for new ideas. Our developing technology and fully open-access content together provide an opportunity to incubate and grow innovation. Ensuring that others can reuse and build on our content also facilitates the long-term preservation of the research output.

PLoS began this process with the PLoS/Mendeley Binary Battle (see Comment 1.c) and will continue it by selecting different early stage innovations to incubate each year, as appropriate. These may take the form of new technologies, services, or partnerships. The goal will be to serve our mission, which will in turn foster new opportunities, growth, and jobs.

Dynamic and creative ideas in many sectors are now regularly emerging from a myriad of startup companies, academic institutions, and industry. By testing and growing these early-stage ideas and technologies, PLoS will contribute to a period of rapid innovation in scientific communication that will both serve science and create economic growth.

4.b Some examples of private–public partnerships include:

(i) Collaborations with new companies and startups:

- Mendeley [9]—late stage startup company gaining market share. PLoS and Mendeley have a joint interest in providing researchers with services and systems for recommending content to each other based on social web analytics. One partnership created was the PLoS/Mendeley Binary Battle (for details see Comment 1.c). The two winning apps were:
    - openSNP uses the PLoS API to allow customers of direct-to-customer genetic tests to publish their test results, find others with similar genetic variations, learn more about

their results, find the latest primary literature on their variations, and help scientists to find new associations [14].
    - PaperCritic uses the PLoS API to offer researchers a way of obtaining and providing feedback for each other's work in a fully open and transparent environment [15]. This is a paper-reviewing service that PLoS can support by offering a plug-in showing reviews while users are reading an article.
- Access Innovations—mature company that PLoS is partnering with to create a Science, Technology, and Medicine taxonomy [32].
- JANE—open–source, student-built tool for "fingerprinting" documents so that you can find related articles [33]. PLoS is helping to grow JANE and make it available more widely.

(ii) Partnerships with researchers still within academia:

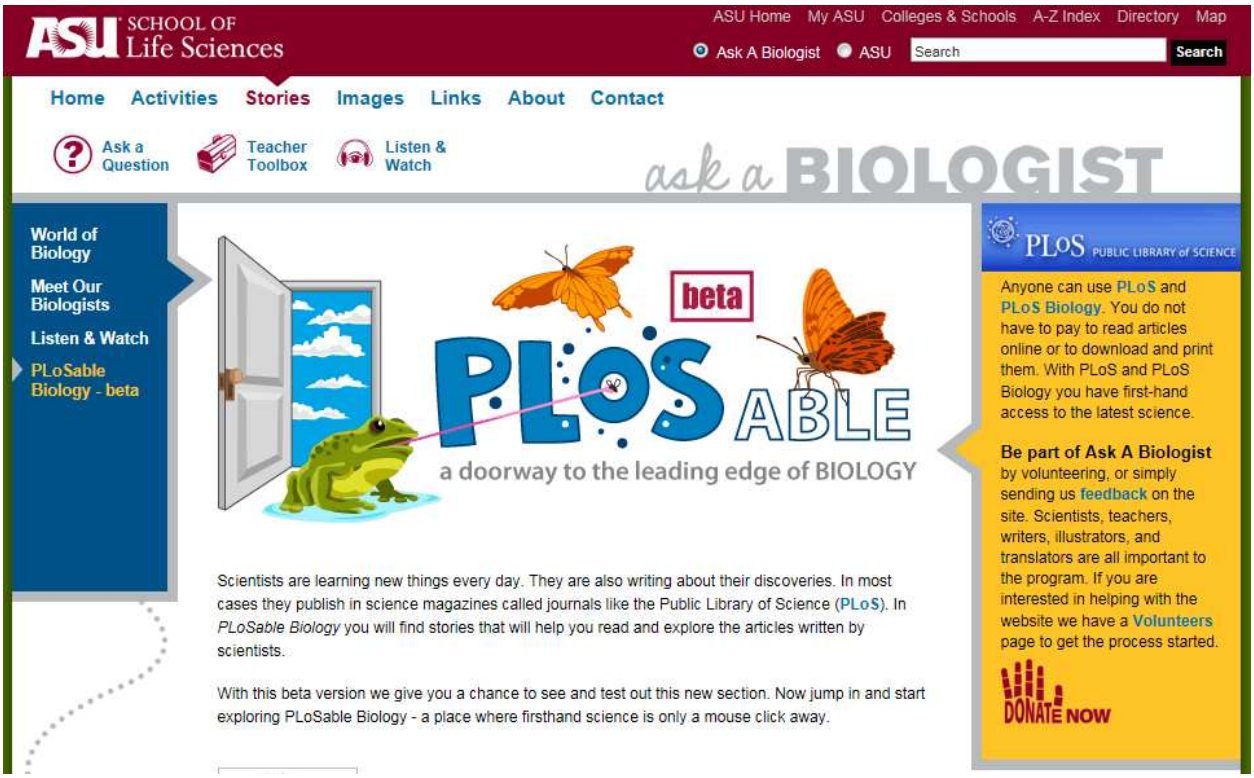
- David Shotton (University of Oxford)—working on Semantic Publishing and Referencing (SPAR) ontologies, has done a pilot in which he enhanced an article from *PLoS Neglected Tropical Diseases* [34] with semantics and rich linking that could be transformed into a web-wide service for publishers that would serve discovery of scientific information [35,36].
- Sri Devabhaktuni (California)—a researcher who built a tool for recommending noun phrases for papers which will aid in discovery in an information overload environment [37].

(iii) Partnerships with established public and private enterprises:

- PLoS collaborated with NCBI and Google's Knol publication platform for PLoS Currents to enable a rapid-response publication service for tracking research about influenza ([20]; see also Comment 1c).
- PLoS partnered with the Census of Marine Life to produce a series of subject-specific open-access collections ([23]; see also Comment 1.c). This is supported by government agencies concerned with science, environment, and fisheries from more than 80 nations, including the US, as well as from private foundations and corporations. Sponsorship was provided by the Sloan Foundation to cover the cost of the publication fees.

(i.v) Partnerships with educational bodies:

- PLoSable Biology is an educational resource featured on Arizona State University's (ASU) "Ask a Biologist" website [38] that grew out of a year-long collaboration between an editor at *PLoS Biology* and Charles Kazilek (Assistant Dean of Technology, Media and Communication at ASU), who runs the ASU site. Because all PLoS content is fully open access and thus erects no barriers to reuse, PLoSable Biology was able to use PLoS research articles and related materials to provide tutorials, stories, and interactive learning opportunities for students (pre-K–12th grade) and their teachers and parents. With a new website feature called PLoSable Biology (beta), "Ask a Biologist" makes selected PLoS articles comprehensible to all through simple summaries that link back to the original article for further reading. The site also provides learning opportunities about the benefits of open-access content as well as a showcase for the public communication of science.



(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?

Comment 5

The metadata of publications enable specific actions to be made to the content, rather than just labeling it. This facilitates the reuse of that content and the multiple components that it comprises (figures, tables, data, key words, semantic mark-up, etc.). It is important that the metadata model supports the appropriate context for the published works with a controlled vocabulary that permits reuse and interoperability between content platforms and databases. It is therefore also important to couple metadata with an API for standards-based data exchange.

Scholarly publishers should, at minimum therefore, support the National Library of Medicine XML DTD standard for content and metadata as well as the Dublin Core. Publishers should also deposit digital object identifiers (DOIs) into repositories such as CrossRef (http://www.crossref.org/) and, similarly, ensure they comply with requirements to allow unique author identification via platforms such as ORCID [39], a central registry of unique identifiers for individual researchers, and an open and transparent linking mechanism between ORCID and other current author identification schemes).

To facilitate new initiatives in the semantic web, publishers should classify content with public domain taxonomies and thesauri and make these classifications available in machine-readable format in the source code. Standard taxonomies and thesauri that are created by, funded by, and recommended by scientific agencies and institutions should, whenever possible, be adopted and used by scholarly publishers to facilitate discovery across the platforms and silos that have been artificially created over the years. These efforts will also enable large-scale research projects across platforms, publishers, and resources.

(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?

Comment 6

The success of any public-access policy will depend on the implementation of consistent requirements from federal funding agencies. If not, the research output of the federal agency that, for example, permits free access but not full open access (i.e., permitting reuse), is likely to lag behind other federal agencies in terms of scientific progress, and knock-on effects such as commercialization, job creation, and economic growth. In addition, many researchers are funded by more than one agency. Consistent access policies and data requirements across federal agencies will therefore avoid duplication of effort and help to minimize costs (e.g., of archiving) while achieving maximum interoperability and visibility of scientific research to the wider community.

Consistent requirements include:

- The same licensing policy permitting unrestricted reuse (CC-BY or equivalent, see Comment 1e).
- Mandating deposition of research publications in publicly accessible repositories (also permitting reuse) so that federally funded research is preserved and available for reuse in perpetuity (see Comment 1d).
- Ensuring there is no delay to full open access of the publication, associated data, and metadata on publication (see Comment 8).
- Encouraging the development of tools and services to evaluate research publications and thus enable third parties to track and analyze research output (for example, PLoS's ALMs initiative, see Comment 1b).
- Encouraging grantees to provide publication management plans alongside data sharing and management plans (see also Comment 8).

(7) Besides scholarly journal articles, should other types of peer-reviewed publications resulting from federally funded research, such as book chapters and conference proceedings, be covered by these public access policies?

The output of all federally funded research should always be made available for reuse under a full open-access license. Although print distribution or author- and reviewer-level payments may incur additional costs, a publication fee paid to the publisher should cover these and guarantee the legal right of free online access and reuse of that material. This should include at least primary research (e.g., original research articles) and secondary research (such as systematic reviews in the medical literature) as well as conference proceedings, book chapters, and scientific protocols that are facilitated by federal funding.

(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?

Comment 8

8.a The appropriate embargo period.

PLoS does not impose any embargo on a published paper because any barrier to the literature means a potential delay to access and reuse of that material by scientists or third parties. As outlined under Question 1, a delay in delivering material to these different audiences not only delays scientific progress, but also hampers innovation, the potential emergence of new markets, economic growth, and jobs. PLoS is evidence that you can have a successful business model without having to impose embargoes. Federal funding agencies should therefore ensure that the entire corpus of research they fund is fully open access, and ultimately they should not permit any delay to the access of published paper. Faster access means faster reuse and commercialization. There are no subject areas or disciplines that should be granted an exemption.

8.b Why there shouldn't be "delayed" open access.

A recent report commissioned by the Research Information Network (RIN), JISC, Research Libraries UK (RLUK), the Publishing Research Consortium (PRC), and the Wellcome Trust [40] concluded that permitting embargoes is unlikely to provide the long-term incentives for publishers to transition to open access and would thus hamper the growth of access:

> *1. The Delayed scenario offers closest to a zero cost. But it depends on voluntary action by publishers, and it is not directly amendable to policy influence … Moreover, it would probably involve embargoes longer than funders such as the Wellcome Trust currently require, it could preclude aggregation of articles in subject repositories, and – as with the Green scenario [self-archiving] – there are risks to the sustainability of the subscription model on which it relies. In our view, therefore, while there is no harm in policy-makers encouraging it as a low-cost and*

> *arguably lower-risk way of expanding access, it is unlikely in practice to provide significant changes in access*.

The Wellcome Trust noted [41] that the report "discusses a number of scenarios and suggests that the "Gold" scenario—the model in which author-side payments are levied to enable immediate open access to the published article—has the potential to achieve the highest benefit-cost ratio while lowering the UK's net costs for scholarly communication. It is also the only model that is considered to be fully sustainable":

> *7. Of the two open access routes, our view is that the Gold route is preferable in the long run, given (i) its underlying sustainability; (ii) the advantages of the author-side business model in terms of improved transparency and lower barriers to market entry, which point to improved economic efficiency; and (iii) (depending on the level of the APC [Article Processing Charge]) the potential to achieve both higher BCRs [benefit-cost ratios] and lower net costs for the UK in general and for its universities in particular.* [40]

Evidence that embargos do not hamper growth and productivity include the faster rate that articles are read, cited, and downloaded [10,11]. Moreover, any delay to accessing the full text of an article can lead to public misinformation by media outlets if third parties, such as the general public or independent researchers, are unable to verify the claims. A notable example is a paper in which NASA claimed to have isolated an "extra-terrestrial" bacterium that substitutes arsenic for phosphorus on its macromolecules and metabolites [42]. Having no embargo, therefore, helps to facilitate the public communication of science. Articles cited by national and international media outlets can provide links to the full article and third parties do not have to pay a fee to independently verify the claims made by journalists.

As the journalist George Monbiot stated recently in the UK national newspaper *The Guardian* [43],

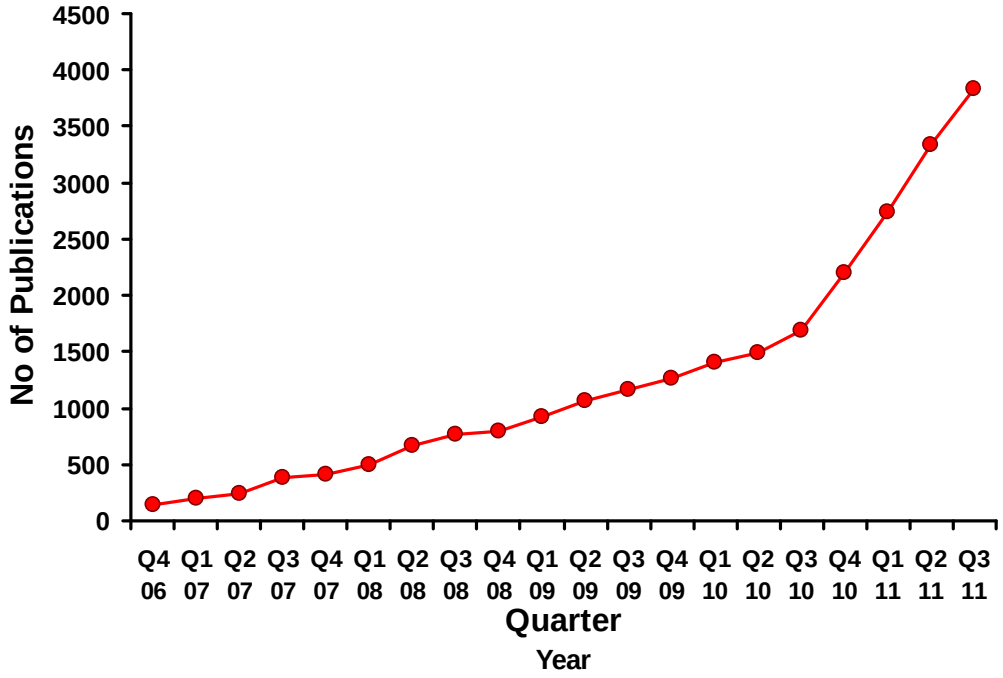
> *I refer readers to peer-reviewed papers, on the principle that claims should be followed to their sources. The readers tell me that they can't afford to judge for themselves whether or not I have represented the research fairly. Independent researchers who try to inform themselves about important scientific issues have to fork out thousands. This is a tax on education, a stifling of the public mind. It appears to contravene the universal declaration of human rights, which says that "everyone has the right freely to … share in scientific advancement and its benefits"*.

It is also worth noting that publishers who were previously resistant to making their content publicly available are now making some of their archives more accessible; for example, the Royal Society in the UK has made their archive of the oldest journal in the world (*Philosophical Transactions of the Royal Society*, launched in 1695) freely available to read [44].

Perhaps the best evidence that embargoes and full open access do not hinder commercialization, profit, or a successful business is the continuing expansion of fully open-access publishers, such as PLoS, BioMedCentral (http://www.biomedcentral.com), Hindawi (http://www.hindawi.com), and Co-Action Publishers (http://www.co-action.net). This is reflected in the almost exponential growth of, for example, publications in the journal *PLoS ONE* (Figure 1, from [45]). This journal has received huge support from authors and the growing editorial board of more than 2,000 academic editors. The publication and editorial model is now being endorsed by other publishers, including publishers that

have traditionally restricted access via a subscription. While some of these publishers are offering Creative Commons licenses that still prohibit commercial reuse (e.g., Nature Publishing Group's *Scientific Reports*), and are therefore not fully open access, they are demonstrating support for a model where the costs of publication are paid up front and there is no embargo on the published paper. In the past year a series of journals have emerged that are very similar to *PLoS ONE* (Table 1, adapted from [45]), suggesting that the landscape of scholarly publishing has irreversibly shifted [45].

Figure 1. Publication growth of *PLoS ONE*.



From [45].

Table 1. A sample of recently launched journals with no embargo and editorially similar to *PLoS ONE*.

Journal Name	**Publisher**	**Website**
G3	Genetics Society of America	http://www.g3journal.org
**BMJ Open*	British Medical Journals publishing group	http://bmjopen.bmj.com
**Scientific Reports*	Nature Publishing Group	http://www.nature.com/srep/
AIP Advances	American Institute of Physics	http://aipadvances.aip.org
**Biology Open*	Company of Biologists	http://bio.biologists.org

TheScientificWorldJournal (TSWJ)	Hindawi	http://www.tswj.com
QScience Connect	Bloomsbury Qatar Foundation	http://www.qscience.com
SAGE Open	SAGE	http://sgo.sagepub.com
SpringerPlus	Springer	http://www.springeropen.com/springerplus/

* These journals have no embargo on access to their articles, but are not fully open access, because they all offer a Creative Commons license that restricts commercial reuse.

Please identify any other items the Task Force might consider for Federal policies related to public access to peer-reviewed scholarly publications resulting from federally supported research.

In addition to ensuring that the peer-reviewed research publications of federally funded research are made fully open access, federal agencies should adopt consistent and parallel policies about the raw data behind any research publication they fund. US and UK funding agencies already appreciate that more open sharing of data is required. As a recent report from the UK House of Commons Select Committee on Science and Technology examining research integrity concluded,

> *Access to data is fundamental if researchers are to reproduce, verify and build on results that are reported in the literature … The presumption must be that, unless there is a strong reason otherwise, data should be fully disclosed and made publicly available. In line with this principle, where possible, data associated with all publicly funded research should be made widely and freely available…The work of researchers who expend time and effort adding value to their data, to make it usable by others, should be acknowledged as a valuable part of their role*. [46]

With some exceptions (such as the deposition of genetic sequence data to GenBank), most research papers provide summary data in the form of figures and tables to support their analyses and conclusions. This does not allow reviewers, readers, or other third parties to extract the data necessary to replicate the analysis, and thereby verify the claims made in the paper. We recommend that federal agencies therefore also:

(i) require that the data needed to replicate a paper is made available for reuse at the same time as the publication of the article.
(ii) require that there is appropriate citation and credit to authors and funders for the data produced.
(iii) help identify appropriate methodologies for citing and sharing data outputs.
(iv) provide incentives for authors by giving credit for good sharing practice.

New initiatives, such the Open Knowledge Foundation's Working Group on Open Data in Science, are already trying to build the infrastructure and culture to facilitate data sharing to this end [47]:

In terms of our primary aim of providing tools, apps, and datasets for generating, discovering, and reusing open data, ideas are flowing continuously but require the input of the wider scientific community in identifying the problems they face in publishing, discovering, and reusing data online and requesting assistance in solving them. The working group aims to provide a community and network that can respond to these needs and a hub for access to the resulting tools, which we hope all stakeholders in scientific data will find valuable. Better science—in terms of transparency, reproducibility, increased efficiency, and ultimately a greater benefit to society—depends on open data.

REFERENCES AND SOURCE MATERIAL

1. Laakso M, Welling P, Bukvova H, Nyman L, Björk B-C, et al. (2011) The development of open access journal publishing from 1993 to 2009. PLoS ONE 6: e20961. doi:10.1371/journal.pone.0020961 Available: http://www.ncbi.nlm.nih.gov/pubmed/21695139. Accessed 22 November 2011.
2. Carroll MW (2011) Why full open access matters. PLoS Biol 9: e1001210. doi:10.1371/journal.pbio.1001210
3. Creative Commons (n.d.) Creative Commons Attribution 3.0 Unported (CC BY 3.0). Available: http://creativecommons.org/licenses/by/3.0/. Accessed 26 Dec 2011.
4. Jensen M (15 June 2007) The new metrics of scholarly authority. The Chronicle of Higher Education. Available: http://chronicle.com/article/The-New-Metrics-of-Scholarly/5449. Accessed 26 Dec 2011.
5. Adler R, Ewing J, Taylor P (2008) Citation statistics: a report from the International Mathematical Union (IMU) in cooperation with the International Council of Industrial and Applied Mathematics (ICIAM) and the Institute of Mathematical Statistics (IMS). Berlin: International Mathematical Union. Joint Committee on Quantitative Assessment of Research.
6. Pendlebury DA (2009) The use and misuse of journal metrics and other citation indicators. Arch Immunol Ther Exp (Warsz) 57: 1-11. Available: http://www.ncbi.nlm.nih.gov/pubmed/19219526. Accessed 26 Dec 2011.
7. Hicks MH-R, Spagat M (2008) The Dirty War Index: a public health and human rights tool for examining and monitoring armed conflict outcomes. PLoS Med 5: e243. doi:10.1371/journal.pmed.0050243. Available: http://dx.doi.org/10.1371/journal.pmed.0050243. Accessed 19 November 2011.
8. Hicks M (2010) First application of the Dirty War Index: CBDARs in Southern Afghanistan. Comment on Hicks MH-R, Spagat M (2008) The Dirty War Index: A Public Health and Human Rights Tool for Examining and Monitoring Armed Conflict Outcomes. PLoS Med 5(12): e243. doi:10.1371/journal.pmed.0050243. Available: http://www.plosmedicine.org/annotation/listThread.action?inReplyTo=info%3Adoi%2F10.1371%2Fannotation%2F3f0bd4ea-9278-4650-938b-5c73b15b07cc&root=info%3Adoi%2F10.1371%2Fannotation%2F3f0bd4ea-9278-4650-938b-5c73b15b07cc. Accessed 22 November 2011.
9. Mendeley Ltd. (2011) Mendeley home page. Available: http://www.mendeley.com/. Accessed 26 December 2011.
10. Bernius S, Hanauske M (2009) Open access to scientific literature - increasing citations as an incentive for authors to make their publications freely accessible. In: IEEE System Sciences, 2009. HICSS '09. 42nd Hawaii International Conference. pp. 1-9. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4755635. Accessed 26 December 2011.

11. The Open Citation Project (2004; updated 2011) The effect of open access and downloads ('hits') on citation impact: a bibliography of studies. Available: http://opcit.eprints.org/oacitation-biblio.html. Accessed 26 December 2011.
12. Brynjolfsson E, Hammerbacher J, Stevens B (2011) Competing through data: three experts offer their game plans. McKinsey Quarterly. Available: http://www.mckinseyquarterly.com/Competing_through_data_Three_experts_offer_their_game_plans_2868. Accessed 26 December 2011.
13. Mendeley Developers Portal (n.d.) Mendeley/PLoS API Binary Battle. Available: http://dev.mendeley.com/api-binary-battle/. Accessed 20 November 2011.
14. openSNP (n.d.) openSNP home page. Available: http://opensnp.org/. Accessed 26 December 2011.
15. PaperCritic (n.d.) PaperCritic home page. Available: http://www.papercritic.com/. Accessed 26 December 2011.
16. Altmetric.com (n.d.) PLoS Explorer. Available: http://altmetric.com/interface/plos.html. Accessed 20 November 2011.
17. Fenner M (n.d.) ScienceCard. Available: http://sciencecard.org/. Accessed 20 November 2011.
18. Taraborelli D (n.d.) ReaderMeter: research impact, crowdsourced. Available: http://readermeter.org/. Accessed 20 November 2011.
19. Public Library of Science [PLoS] (n.d.) PLoS Currents: Influenza. Available: http://knol.google.com/k/plos-currents-influenza#. Accessed 26 December 2011.
20. Varmus H (2009) A new website for the rapid sharing of influenza research. The official PLoS Blog. Available: http://blogs.plos.org/plos/2009/08/a-new-website-for-the-rapid-sharing-of-influenza-research/.
21. PLoS (n.d.) PLoS Currents: Disasters. Available: http://currents.plos.org/disasters/. Accessed 26 December 2011.
22. PLoS (n.d.) PLoS Collections. Available: http://www.ploscollections.org/. Accessed 20 November 2011.
23. PLoS (n.d.) Census of Marine Life Collections. Available: http://www.ploscollections.org/static/comlCollections.action. Accessed 26 December 2011.
24. PLoS (2011) PLoS Biodiversity Hub. Available: http://hubs.plos.org/web/biodiversity/. Accessed 20 November 2011.
25. Houghton J, Rasmussen B, Sheehan P (2010) Economic and social returns on investment in open archiving publicly funded research outputs Report to SPARC. Centre for Strategic Economic Studies Victoria University. 41 p. Available: http://www.arl.org/sparc/bm~doc/vufrpaa.pdf. Accessed 9 January 2012.
26. Houghton J, 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). Available: http://www.jisc.ac.uk/publications/reports/2009/economicpublishingmodelsfinalreport.aspx. Accessed 11 January 2012.
27. Houghton J, Sheehan P (2006) The economic impact of enhanced access to research findings. CSES working paper no. 23. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.136.1066. Accessed 9 January 2012.
28. Wellcome Trust (2003) Economic analysis of scientific research publishing. A report commissioned by the Wellcome Trust. Available: http://www.wellcome.ac.uk/About-us/Publications/Reports/Biomedical-science/WTD003181.htm. Accessed 26 December 2006.
29. Mcguigan GS, Russell RD (2008) The business of academic publishing: a strategic analysis of the academic journal publishing industry and its impact on the future of scholarly publishing. Electronic Journal of Academic and Special Librarianship 9. Available:

http://southernlibrarianship.icaap.org/content/v09n03/mcguigan_g01.html. Accessed 9 January 2012.
30. Carroll MW (2011) Why full open access matters. PLoS Biol 9: e1001210. doi:10.1371/journal.pbio.1001210. Available: http://dx.doi.org/10.1371/journal.pbio.1001210. Accessed 26 December 2011.
31. OpenAIRE (n.d.) Paving the way to an open scientific information space: OpenAIREplus – linking peer-reviewed literature to associated data. Available: http://www.openaire.eu/home/76-highlights/326-openaireplus-press-release?lang=en. Accessed 6 January 2012.
32. Access Innovations, Inc. (n.d.). Access Innovations home page. Available: http://www.accessinn.com/. Accessed 26 December 2011.
33. Journal/Author Name Estimator (n.d.) Journal/Author Name Estimator. Available: http://biosemantics.org/jane/. Accessed 26 December 2011.
34. Reis RB, Ribeiro GS, Felzemburgh RDM, Santana FS, Mohr S, et al. (2008) Impact of environment and social gradient on Leptospira infection in urban slums. PLoS Negl Trop Dis 2: e228. doi:10.1371/journal.pntd.0000228. Available: http://dx.plos.org/10.1371/journal.pntd.0000228. Accessed 22 December 2011.
35. Shotton D, Portwin K, Klyne G, Miles A (2009) Adventures in semantic publishing: exemplar semantic enhancements of a research article. PLoS Comput Biol 5: e1000361. doi:10.1371/journal.pcbi.1000361. Available: http://dx.doi.org/10.1371/journal.pcbi.1000361. Accessed 22 December 2011.
36. Shotton D (2009) Semantic publishing: the coming revolution in scientific journal publishing. Learned Publishing 22: 85-94. Available: http://openurl.ingenta.com/content/xref?genre=article&issn=0953-1513&volume=22&issue=2&spage=85. Accessed 22 December 2011.
37. Srikrishna D, Coram MA (2011) Using noun phrases for navigating biomedical literature on PubMed: how many updates are we losing track of? PLoS ONE 6: e24920. doi:10.1371/journal.pone.0024920. Available: http://dx.doi.org/10.1371/journal.pone.0024920. Accessed 22 December 2011.
38. Arizona State University School of Life Sciences (n.d.) Ask a biologist. PLoSable Biology – beta. Available: http://askabiologist.asu.edu/explore/plosable. Accessed 26 December 2011.
39. Open Researcher & Contributor ID [ORCID] (n.d.) ORCID home page. Available: http://orcid.org/. Accessed 26 December 2011.
40. CEPA LLP (Joel Cook, Daniel Hulls and David Jones) and Mark Ware Consulting Ltd (Mark Ware) (2011) Heading for the open road: costs and benefits of transitions in scholarly communications. Report Commissioned by the RIN, Research Libraries UK, the Wellcome Trust, the Publishing Research Consortium and the Joint Information Systems Committee. Available: http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/heading-open-road-costs-and-benefits-transitions-s. Accessed 23 December 2011.
41. Wellcome Trust (7 April 2011) Heading for the Open Road: costs and benefits of transitions in scholarly communications [press release]. Available: http://www.wellcome.ac.uk/News/2011/News/WTVM050833.htm. Accessed 26 December 2011.
42. Zimmer C (27 May 2011) The Discovery of arsenic-based twitter. Slate. Available: http://www.slate.com/articles/health_and_science/science/2011/05/the_discovery_of_arsenicbased_twitter.html. Accessed 26 December 2011.
43. Monbiot G (29 August 2011) Academic publishers make Murdoch look like a socialist. The Guardian. Available: http://www.guardian.co.uk/commentisfree/2011/aug/29/academic-publishers-murdoch-socialist. Accessed 23 December 2011.

44. The Royal Society (26 October 2011) Royal Society journal archive made permanently free to access [press release]. Available: http://royalsociety.org/news/Royal-Society-journal-archive-made-permanently-free-to-access/. Accessed 23 December 2011.
45. MacCallum CJ (2011) Why ONE is more than 5. PLoS Biol 9: e1001235. doi:10.1371/journal.pbio.1001235. Available: http://dx.doi.org/10.1371/journal.pbio.1001235. Accessed 26 December 2011.
46. House of Commons Science and Technology Committe (2011) Science and Technology Committee – eighth report. Peer review in scientific publications. Available: http://www.publications.parliament.uk/pa/cm201012/cmselect/cmsctech/856/85602.htm. Accessed 23 December 2011.
47. Molloy JC (2011) The Open Knowledge Foundation: open data means better science. PLoS Biol 9: e1001195. doi:10.1371/journal.pbio.1001195. Available: http://dx.doi.org/10.1371/journal.pbio.1001195. Accessed 23 December 2011.