

Request for Information on Public Access

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?

a. Growing Markets

The growth of existing markets and the development of new markets will both be accomplished most successfully by the provision of immediate, full access to and reuse of complete collections without commercial restrictions. This complete access will permit entrepreneurial members of the public to fully use these works to generate new services and products unencumbered by restrictions that might limit their innovative use. The sooner and more completely these works are available the more quickly and fully individuals and companies will be able to unlock their economic potential.

In particular, open works are more likely to be effectively commercialized because businesses will be able to operate without major restrictions that limit their creative application. Access based on limitations creates disincentives for use generally and may curtail the unexpected or inventive uses that a particular company might commercialize.

Where complete access is not made available companies will be forced to choose between a presumption of limited use or reliance on copyright exceptions such as fair use. As noted above, limited use reduces the ability of all users to pioneer groundbreaking applications and may make many potential investors hesitant to use works that carry such a limitation.

Though copyright exceptions play an important role in unlocking these works, statutory exceptions are generally aimed at existing communities and may not protect the innovative uses that supports the expansion of new and developing markets. Similarly, although fair use is a vital tool for users, those seeking to commercialize these works are likely to find the uncertainty inherent in this exception a significant deterrent to investment and entrepreneurship. Without

open access reliance on fair use would force a company into a legally uncertain position likely to make investors uncomfortable and force repeated legal analysis at significant cost.

Truly open public access also empowers more users and particularly new users to keep abreast of the latest trends in the research. Greater dissemination of the latest research can be expected to support greater innovation based on that research. Even in cases where the specific content is not applied in a market context expanding the base of knowledge will improve the intellectual marketplace and raise the quality of research across the board. Unexpected users in particular will be empowered to fully-engage with this content so as to discover commercial applications that would otherwise be missed.

Faster – ideally immediate – access is just as important. The sooner work is available to the public the more quickly all citizens will be able to apply ideas generated by the research thus leading to new products and services entering the marketplace more quickly.

Public access drives new industries and faster access facilitates new jobs across all segments of the economy. Knowledge-rich professions such as agricultural and biotechnical sciences, high tech professions such as energy, and information professions such as publishing all rely heavily on the sort of research at issue here and all of these professions are major source of sustainable, high-quality American jobs.

Access to this information will also incentivize private investment in technical solutions that build on government research – a traditional strength of the American economy that is badly needed today. We already have examples of IT infrastructure that aggregates and mines public information research such as Google Scholar and goPubMed. With full, immediate public access these companies can offer better services with increased commercial potential. New jobs and new companies can be developed to further capitalize on work that the government is already paying for.

b. Driver of Scientific Productivity

Scientific innovation and productivity also rely on open access to research funded by the federal government. We already have strong empirical evidence that open access research is read by more people. This means that open access research promotes more and faster follow-on

research as scientists use these works in innovative ways. Open access research also encourages a greater diversity of follow-on research as many minds explore different and unexpected angles, including research pathways that might otherwise have been missed. By letting all scientists incorporate the results of governmentally-funded research into their own work more quickly, open access will encourage faster, more thoughtful application of that research towards the next generation of scientific innovation.

Open access content is equally valuable for use by new scientific tools such as machine-readers and computational analysis. Fast access to all data is necessary for scientists to leverage these new tools so they can identify better content and scientific research can progress more quickly, and more intelligently.

Machine-reading in particular opens up entirely new scientific pathways, enabling the discovery of new connections across the body of research. This powerful new tool, however, relies on complete access to large bodies of data with no limitations. With complete open access new research pathways and semantic tools can be used to speed and transform scientific productivity. This, in turn, opens new avenues for commercial development that capitalizes on existing public investments.

Finally, open access permits unforeseen participants to join in the scientific enterprise. American history is filled with scientists and technological innovators who were not affiliated with established institutions and the recent rise of internet success stories in particular epitomizes the value of the unexpected innovator. Members of the public who might not otherwise have access to this research will be able to contribute in the tradition of amateur innovators such as Steve Wozniack, Bill Gates, and Mark Zuckerberg.

In the academy and the research laboratory, traditional scholars in related disciplines will also be empowered by open access to contribute to scientific progress across all disciplines. This paves the way for innovated interdisciplinary discoveries. It also increases the return on investment for all research since it will be able to be used across all contexts. By opening access to all citizens scientific progress can be driven by thinkers across disciplinary boundaries and beyond the walls of traditional scholarly institution to harness the American innovative spirit.

c. Costs and Benefits

i. Benefits: The benefits of open access have been demonstrated by several major governmental programs that are already in effect. The Houghton Reports on FRPAA make it clear that opening up access produced at least a fivefold increase in return on investment. The benefits of an open access policy similar to that of the existing NIH policy are estimated at approximately 8 times larger than the costs. The net present value gains of expanding an NIH-style policy to all other U.S science agencies is estimated to be on order of \$1.5 billion (net costs of running the archive). Of that figure, approximately 60% is estimated to accrue directly to the U.S. economy.

Open access provides the additional benefit of providing increased accountability for federal agencies. Outcomes of funded research will be easier to measure and Congressional budget drafters, appropriators, and authorizers will have better information to assess the value of existing expenditures and target funding on the most promising research. Policymakers will also have better information across the board based on the improved access and use of research.

ii. Cost: The NIH's open access policy provides a closely analogous example that illustrates the cost-effectiveness of open access policies. The NIH has proved cost-effective with between \$3.5 and \$4.6 million – or about 1/100th of 1 percent of the NIH's \$30 billion budget – providing access to better than 2.2 million articles. These articles are used by more than 500,000 users per day, most of whom come from outside of the traditional university environment. There is deep demand for this information across the public sector.

This use of NIH content underscores the cost-effectiveness of open access and provides an important base for expanding open access. By building on these existing programs existing infrastructure can be leveraged to avoid duplicative effort. This base can then be used to expand access to additional content at a minimal incremental cost.

d. Type of Access Needed

Free, immediate access that includes the right to reuse will have the greatest benefit for scientific progress, technological innovation and the American economy. Any restrictions on access to the material paid for by the public will limit the value of that information and significantly diminish the return on the public's investment. Full reuse will permit researchers to maximize the value of this work as well as unlocking additional value in the years to come. It

will also limit duplicative costs and build on the results of this research in sustainable ways that will continue to sustain scientific progress and commercial innovation for decades.

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?

Open access fits comfortably within the current copyright regime which balances the right of the public to use works and the intellectual property rights of the authors that create and the agencies that fund their work. Successful open access systems such as the NIH permit use of works based on existing copyright mechanisms such as fair use and the eventual entry of works into the public domain.

Along with these established copyright rules, greater utility should be enabled to permit use that supports the scientific and commercial innovation that public funding is designed to encourage. Mechanisms to enable full use of this material should be included in the policy so that users can engage in distribution, reuse, text mining, data mining, computation, and the creation of derivative works.

Adoption of a licensing system similar to the Create Commons “CC-BY” license will permit full use of this content by the public even in cases where existing copyright exceptions are not available. Licensing mechanisms such as the Creative Commons have been extraordinarily effective because they are simple to create and attach, easy for users to understand, and fit comfortable with the open use of content created to serve the public good.

An IP regime that balances the rights of all parties will best-serve the public. An embargo period will permit publishers to commercialize these works while users are able to rely on fair use for the comment and criticism needed to evaluate new articles. After this embargo period works should be available to the public subject to a standard CC-BY license that assures attribution but otherwise unchains works so that the public can maximize their value.

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?

The federal government is the appropriate entity to provide permanent stewardship of these articles, and is in a unique position to ensure that publicly funded articles are permanently preserved, made accessible, and useable. As such, any public access policy must give the government the rights to archive and distribute these works.

At a minimum, the government must maintain an accessible, mirrored version of all content so that the public can be assured of having access. We have numerous examples from other agencies such as the SEC and USPTO of the federal government maintaining large databases of information. The closest analogy, the NIH, has proven to be extremely cost-effective: NLM reports PMC costs less than 1/100th of one percent of NIH' s operating budget to run.

Distributions across multiple repositories is not a problem but all repositories must have the same conditions surrounding access and use to ensure genuine long-term storage and sufficient interoperability. Repositories that meet conditions for public accessibility, use rights, interoperability and long-term preservation of articles, could be maintained by third parties. This would encourage innovative public/private partnerships and permit numerous companies to develop tools and search strategies that improve search efficiency much as companies such as Westlaw and Lexis generate millions of dollars every year by supporting access to legal documents in the public domain.

A “dark” archive that does not provide access to all parties is not an acceptable solution. Efforts to archive content must be measured in decades, not years, and library experiences have shown that regular access/use of digital materials is crucial element in effective long-term preservation. Without regular access/use, archival veracity cannot be ensured and public access may be limited by whatever institutions do make the content available if they push the boundaries of accessibility based on format, etc.

The federal government making this content available is not duplicative; it is necessary to ensure this public investment is protected and fully-leveraged. Current market attempts at archives are not adequate. For example, Cornell and Columbia report that only ~15% of their combined journal holdings are currently archived by LOCKSS and Portico combined.

Whether a centralized or decentralized model is chosen, all works must be made available to the public in such a way that preservation, access, and use are fully protected. The federal government has the infrastructure and the mandate to do this.

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?

The most successful models will be those that recognize all of the partners in the research enterprise. Publishers, libraries, intergovernmental organizations, scientific communities across national borders and of course scholars themselves all have a stake in this process. A narrow focus on “existing publishers” risks missing the important contributions of these other stakeholders as well as the invaluable advances made by the next generation of innovators.

Publishers – both established and forthcoming – can play an important role in this process by providing approved repositories that meet conditions for public accessibility, use rights, interoperability and long-term preservation of publicly funded articles. No single stakeholder, however, should be given a monopoly on these works financed with public dollars. Partnerships should permit multiple points of access for users and must be open to anyone in the marketplace who can improve on existing services or offer competing models for innovative use.

Public-private partnerships with academic stakeholders are another important opportunity that should not be ignored. Universities and libraries have extensive experience and existing archive infrastructure, and should be actively encouraged to partner with federal agencies. Empirically, None of the 50+ research funders who currently have public access policies are using publisher sites as the final archives. There are, however, good examples of funders partnering with academic and research institutions in this role.

We have several examples to draw on in this area. In Europe, the Digital Repository Infrastructure Vision for European Research (DRIVER) provides a test case for interoperation of both data network and knowledge repositories as integral parts of the E-infrastructure for research and education on a scale comparable to the United States.

At North Carolina State University we have had success with our own repository, as have our colleagues at partner institutions in the Research Triangle, an area where the broad dissemination of scientific information has led to internationally-praised innovation and substantially boosted the economy. We have also had success with larger partnerships such as the Hathi Trust and Open Library Initiative. As we have seen at NCSU, as well as through initiatives such as Mendeley, PLoS, and even Google, the most successful partnerships are those that recognize all stakeholders in the research enterprise, as well as the public good that they ultimately serve.

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?

Policy surrounding metadata should recognize that metadata is more than a simple description of an item, it is a means for enabling specific actions. As such, metadata should be designed to facilitate specific, desirable actions around use, reuse, and analysis of published works. To enable this use metadata should be machine-readable, particularly for use and reuse.

Creation of metadata should begin with existing standards. Standards such as Dublin Core for exchange, ORCID1/2 for controlled identifiers, and Counter/Sushi for usage tracking provide a base of established and tested models that can be evaluated and improved upon as time passes. Established metadata agencies such as NICO and LOC that have spent years developing

expertise on metadata interoperability should be involved in the ongoing development of new standards.

It is important that metadata be coupled with an API for standards-based data exchange. Published articles and data are distinct issues and metadata must be cognizant of these differences, particularly since articles can also be used as data. Metadata can be used to build bridges between these two through semantic relationships, unique identifiers, and similar coding. The most successful metadata will build on existing standards to enable the specific actions required to maximize access, use, and archiving of these important public 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?

In order to minimize the costs open access policies must be based on consistency of requirements and mandates are essential across disciplines. Researchers often hold grants from multiple agencies and consistent policies will reduce inefficiencies for institutions and individuals. Specifically, open access policies should include uniform requirements for peer-reviewed literature, uniform deposit requirements that reduce complexity and cost. Uniformity can also be expected to increase compliance.

Maximizing the return on taxpayer investment can be accomplished through several principles. First, the policy should take advantage of existing protocols to make deposit in multiple repositories as efficient as possible. This can be accomplished with tools such as SWORD and additional tools that should be developed at the encouragement of the policy.

Articles should also be integrated with grants management systems. This will increase efficiency as well as agency accountability. Properly run, this open access system can be an important tool for providing better information to taxpayers about what they are getting for their investment.

Public access policies also offer an opportunity to enhance productivity management tools in the academy. Universities will be able to better measure research output. They can facilitate the creation of better bibliographies and PI tools, and universities/libraries to use repositories as teaching tools (i.e., teaching scholars more effective literature analytics, etc.

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?

All scholarly and educational materials created with taxpayer funding should be made readily available to the public. Comparable efficiencies support wide dissemination and similar benefits can be expected to accrue to the academic enterprise, scientific innovation and the American economy when these materials are made available.

It should be recognized, however, that different issues arise with different types of material. As such, policies for distinct materials may reflect the distinct nature of those materials. There are important differences between the ecology of journal articles, book chapters, and other educational material and these differences may require different policies. For example, text books are designed for a specific audience

Similarly, different types of educational material are created in the context of different existing models. As such, the policies for different materials may have to be adjusted so as to minimize disruption of those distinct models. The incentive structure for journal articles is built on reputation and prestige but not on financial rewards. Monographs, on the other hand, may be created with less concern about the reputation of the publisher and more focus on financial rewards.

Public access policies should be cognizant of the differences between different types of educational materials but not at the cost of core principles such as openness, access, and maximized efficiency of taxpayer resources.

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?

Access that is complete and immediate best-serves the interests of the American public. The more quickly and readily works are made available the better-able all citizens will be to maximize the scientific and commercial potential of taxpayer-funded works.

If the decision is made to reduce the benefit to the public in order to support the current subscription model used by some academic publishers an author-defined embargo of no more than 12 months could represent an acceptable compromise. An embargo determined by the author of the work that runs between 6 and 12 months would permit publishers to commercialize works at the height of their value while still permitting relative quick access by the public so that these works can be used to grow the economy and drive innovation.

This 0-12 month embargo period has been used across most major disciplines with great success. It represents the norm for the industry and has been adopted by hundreds of journals. Despite concerns when embargoes were first adopted, no one has presented any data demonstrating that this policy has harmed publishers.

Indeed, early concerns about openness are increasingly being replaced by groups such as the Royal Society embracing open access. Royal Society, publisher of the world's first scholarly journal, recently opened access to their back file of articles with a 12 month embargo period, noting that this prestigious and heavily cited back file, dating back to 1665, accounted for less one half of one percent of their overall publishing revenue.

If an embargo is employed then calculation of the effect of the embargo must consider all factors. The assumption that access – embargoed or otherwise – reduces profits for publishers cannot be accepted uncritically. Numerous market conditions interact to generate effects in subscription rates. Growth of journals and papers in disciplines, the price – and pricing history - of a given journal and of competitive titles, the potential impact of required bundles, larger

library budget numbers and trends, and real revenue resulting from “long-tail” business all play an important role.

All of these market conditions regularly contribute to journal cancellations and must be accounted for so that effect of embargo period can be adequately isolated. The most successful embargos will be brief – lasting only as long as a critical evaluation of all market factors can justify – and in line with the established model that does not exceed 12 months.