

**Federally Funded Research:
Examining Public Access and Scholarly Publication Interests**

**Written testimony submitted to the
House Committee on Science, Space, and Technology,
Subcommittee on Investigations and Oversight**

**March 29, 2012
Rayburn House Office Building, Room 2318
10 a.m. EDT**

**H. Frederick Dylla, Ph.D.
Executive Director and CEO
American Institute of Physics**

Executive Summary

The American Institute of Physics (AIP) appreciates the opportunity to submit testimony as part of the House Committee on Science, Space and Technology, Subcommittee on Investigations and Oversight hearing on “Federally Funded Research: Examining Public Access and Scholarly Publication Interests.”

AIP strongly supports the broad dissemination of scholarly research, which includes public access to journal articles, data, and related information. It is our position that the America COMPETES Reauthorization Act of 2010 (P.L. 111-358) contains the best and most effective framework to broaden public access. Indeed, recent collaborative efforts by AIP, other scientific societies, the White House Office of Science and Technology Policy, the Department of Energy, and the National Science Foundation have made significant progress toward these goals under the comprehensive structure outlined in America COMPETES.

Due to the complex and highly nuanced nature of this issue, new legislation or government policies that force free public access through a single dissemination business model would harm the collaborative process that stakeholders have worked so hard to achieve. Moreover, a blanket approach to public access would diminish the quality and value of published scholarly research and actually detract from achieving the goal of increasing access to scholarly literature.

Notes on Written Testimony Structure: The bulk of this written testimony is attached in the form of AIP's December 2011 responses to two Requests for Information (RFIs) from the White House Office of Science and Technology Policy (OSTP), which capture comprehensively AIP's positions on the key policy and technical issues in the public access debate.

About AIP and Issue Relevance

AIP is a 501(c)(3) not-for-profit membership corporation founded in 1931 for the purpose of promoting the advancement and diffusion of the knowledge of physics and its application to human welfare. AIP collaborates with its ten Member Societies to provide resources for activities such as scholarly publishing and outreach to the science community and the general public. Publishing scientific journals is the primary means by which scientific societies communicate advances in research to the community. Publishing is also AIP's primary source of revenue, supporting its outreach activities, which serve the broad physics community and the general public.

AIP is an umbrella organization of ten Member Societies that collectively represent a broad cross-section of more than 135,000 scientists, engineers, and educators in the global physical science community. With an extensive catalog of top-cited journals, AIP is one of the world's leading publishers in the physical sciences. AIP publishes 13 journals; two magazines, including its flagship publication *Physics Today*; and the AIP Conference Proceedings series. In keeping with its goal to increase access to and use of its journals, AIP reinvests its journal revenue in innovative electronic publishing technologies for scholarly journals and offers full-solution publishing services for many of its Member Societies.

In 2011, AIP published more than 15,000 scholarly articles in its journals. AIP also published more than 5,000 additional articles for its Member Societies for which it is the publisher of record (American Association of Physics Teachers; Acoustical Society of America; American Association of Physicists in Medicine; AVS: Science & Technology of Materials, Interfaces, and Processing; and The Society of Rheology).

Four AIP Member Societies (American Astronomical Society, American Geophysical Union, American Physical Society, and The Optical Society) manage their own publications, which collectively published an additional 39,000 articles in 2011. AIP and its Member Societies published approximately one quarter of the articles in the discipline of physics (approximately 240,000) in 2011.

AIP's journal revenues finance its entire publishing process, which includes highly skilled editorial management employing over 130 international scientific editors; end-to-end manuscript oversight with authors; coordination of the essential peer review process (which ensures scientific integrity); translation of the text and figures into the form visible on the published page; final publication (both in the traditional print journal and online and mobile versions); the development, dissemination, and maintenance of searchable and accessible journal databases; and preservation of the digital version throughout the coming decades. This enterprise requires extensive human and capital resources: for its archival journals, AIP expends more than \$40 million annually for the entire operation, including editorial, production, bibliographic tagging, printing, online-hosting, and archiving tasks. AIP employs more than 200 people at our facilities in Melville, N.Y., and College Park, Md., who manage and support our publishing operations.

AIP uses the net revenues from its journal publishing operations to support its Physics Resources Center. This Center provides a variety of outreach services for the scientific community and the general public, including: media services for translating summaries of journal articles for dissemination in lay language media channels; operation of the Niels Bohr Library and Archive for preserving the history of the physical sciences; the Statistical Research Center, which tracks education and workforce statistics for physical scientists; a comprehensive science news service for educating the general public about advances in research; and administration of the Society of Physics Students on more than 700 university and college campuses [see AIP's Annual Reports for details: <http://www.aip.org/aip/reports.html>]. The \$21 million in operating expenses to run these programs is partially offset by net revenues from AIP's journal publishing operations.

About Dr. Dylla

Since 2007, H. Frederick Dylla has served as Executive Director and CEO of AIP. Previously, Dylla served as Chief Technology Officer and as Associate Director at the Department of Energy's Jefferson Lab, where he spearheaded the Free Electron Laser (FEL) program. From 1975 to 1990, he held various positions at the Department of Energy's Princeton Plasma Physics Laboratory, where he helped develop technology for nuclear fusion reactors, particle accelerators, and materials processing for the microelectronics industry. He received his Ph.D. in physics from MIT (1975), is a Past President and Fellow of the American Vacuum Society

(AVS), and a Fellow of the American Physical Society and the American Association for the Advancement of Science. During his scientific career, Dylla has been an author of more than 190 peer-reviewed publications in scholarly literature and has served as an editor and on the editorial board of several journals and monograph series. He presently serves on the boards of the Professional and Scholarly Publishing Division of the Association of American Publishers and the International Association of Scientific, Medical and Technical Publishers. In 2009, Dylla helped organize and participated in the Scholarly Publishing Roundtable under the aegis of the U.S. House Science and Technology Committee. The Roundtable developed consensus recommendations for the development of public access policies for scholarly data and publications.

Access to Journal Content

Scholars and research scientists access AIP published content in very large numbers through several different channels, including subscription or license fees which are generally paid by their institutional libraries. AIP's subscription prices are competitively priced within the physics publishing market. With more than 13 million full-text downloads in 2011, the subscription-based cost per download for AIP content is in the \$2-3 range, representing outstanding value for journals that are in the top rank of their class in terms of scholarly impact. AIP has also responded to budget challenges faced by the library market by offering other access models such as a low-price article rental program that has since been adopted by more than 40 other scholarly publishers. Overall, AIP provides a number of cost-effective and efficient means to access high quality peer reviewed content. Some of these access options are detailed further in the attachments.

AIP has joined a diverse group of journal publishers that make their articles freely available to academics and others in 100 developing countries. Some well-known programs include the United Nation's HINARI, AGORA, and OARE Research4Life programs, HighWire's Developing Economies Program, and JSTOR's Developing Nations Initiative. Additional programs include those of EIFL, INASP, and TEEAL. For descriptions of these and more, see www.library.yale.edu/~llicense/develop.shtml.

Notably, one of AIP's Member Societies, the American Physical Society, spearheaded an initiative that allows public access to all of their journals by making them available at no charge through public libraries and high schools around the country. So far, more than 600 libraries have signed up for this service.

Many librarians have become advocates of open access in response to cost pressures induced by the rapid growth in journals, proliferation of new journals in niche subjects, and the high relative price of some journals. Since library subscribers are AIP's most important customers, AIP and its Member Societies are very sensitive to their concerns and believe that the inclusive framework established under the COMPETES law can help address their concerns without threatening the quality and essential services to the scholarly community provided by scholarly publishing.

Rapidly Changing and Vibrant Marketplace

Within the scholarly publishing realm, new publishers, journals, and business models are continually emerging, signaling a healthy, competitive marketplace. It is AIP's belief that the government should support and encourage this diversity through its actions and policies via mutually beneficial partnerships with publishers, which would contribute to the U.S. economy and maximize the productivity of the scientific enterprise. This ability of scientific publishers to experiment with different publication, business, and access models is essential and assures the vitality, diversity, and effectiveness of the scholarly communication marketplace, leading to scientific and technological advances. This tradition of innovation in communications in the free market is a hallmark of the intersection of research, entrepreneurialism, and publishing going back to the earliest days of our nation.

Freely Available, But Not Free of Cost

AIP understands the enthusiasm for open access for the obvious reason that it increases access to research, which is at the core of our mission. AIP has been trying to build awareness among all affected parties that while open access may mean freely available, the costs to assure the quality, rigor, discovery, and production value of scientific publishing are not zero.

As Maria Leptin, the Director of the European Molecular Biology Organization, wrote in a March 16, 2012 editorial in the journal *Science*: *"Any transition to open access on a large scale will require a clear understanding of the financial challenges that will be faced. Put simply, publishing costs money, and open access does not mean 'for free' – someone must foot the bill."*

Currently, more than 25,000 scholarly journals are being published worldwide, and institutional subscriptions generate income for 90 percent of these titles. For most of the remaining 10 percent, authors or sponsoring agencies pay an upfront fee per article. These articles are posted on the web without subscription barriers as soon as they are published.

Potential Legislative and Policy Impacts

Current policy efforts to increase public access have focused on two approaches: (a) accelerate the transition from the subscription model to the open access model or (b) mandate the release of subscription content after a specified embargo period. The NIH Public Access Mandate, which was introduced in 2008, requires scholarly articles to be posted on NIH's PubMedCentral website 12 months after publication, if any of the authors had NIH funding for any portion of the underlying research reported on in the articles. The proposed Federal Research Public Access Act (FRPAA), introduced in both houses of Congress last month would extend the release mandate to all disciplines represented by the eleven federal agencies that fund research and decrease the timeframe to 6 months.

It is AIP's position that neither of these one-size-fits-all approaches is an appropriate solution for the diverse array of journals published across all the disciplines represented by federally funded research efforts. The open access model is growing at a reasonable rate for fields where such a model is appropriate (i.e., well-funded or fast-moving disciplines). Delayed-release models are not viable for fields where articles have citation lifetimes of years, such as mathematics, theoretical physics, and the social sciences. Additionally, the scholarly community

should determine the methods of dissemination that are appropriate to their fields. Forcing the adoption of either model would likely cause significant harm to the enterprise of scholarly publishing. Furthermore, forcing the adoption of these models is not necessary, given the natural pressures of the marketplace that continually drive the industry to evolve and innovate a wide array of products and dissemination methods.

Agency/Publisher Pilot Projects Launched

The most appropriate role for the federal government is to encourage federal agency/publisher partnerships, examples of which have arisen as a direct result of COMPETES. AIP has been a leading participant in organizing working groups that are proposing and planning partnerships with NSF and DOE on access, linking of grantee reports to publications, data mining across agency/publisher databases, tools and methodology for identifying publicly funded work, and potential pilot projects in the above areas.

Specifically, the DOE Office of Scientific and Technical Information (OSTI) is collaborating with journal publishers to improve DOE's ability to demonstrate the outcomes of the research it funds. This involves engaging with publishers to identify and broaden access to the journal articles reporting on research funded by DOE. To this end, OSTI has embarked on a pilot project to enhance journal article full-text searching, with the intent to make citations of DOE-funded journal articles available in the search and retrieval applications operated by OSTI.

In this pilot project, the scholarly publisher Wiley provides citations to OSTI, including abstracts and hyperlinks to a landing page for the publisher's version of the article. Wiley provides the full text of the article for use in OSTI's archive, which improves search precision and recall. Through this existing infrastructure OSTI would make the journal publisher's full text searchable.

One of the most important tasks that the publishing community has undertaken (funded by subscription and other revenue) is ensuring the proper tagging of articles, verification and disaggregation of author names, and references. An initiative in this vein is the FundRef project—a collaboration involving CrossRef (a Digital Object Identifier registration agency), several funding agencies and publishers to establish protocols for identifying the funding agencies associated with journal articles. OSTI is working with CENDI (a federal scientific and technical information managers group) members to have a standard list of agencies so there will be no ambiguity for agency names.

CrossRef has agreed to add two new metadata elements to the CrossRef database, which would apply to each journal article: funding agency identity and grant number. FundRef pilot demonstrations with at least five major publishers including AIP will be deployed next year.

Similarly, officials at NSF are assuming a leadership role in initiating one-to-two-year pilot projects on expanded public access to research results; these involve universal identifiers for better search results and linking between NSF and publisher databases.

The NSF Directorate for Math and Physical Sciences is in discussions with AIP and the American Astronomical Society to establish a pilot project to link the data behind figures and tables with publications.

Based on these valuable experiences, other agencies and publishers can gain valuable insight on how future collaborations might be structured to promulgate the success of these initial agency-publisher partnerships.

The America COMPETES Act requires the U.S. federal agencies that fund scientific research to develop policies for access to and interoperability among databases, and archiving for data and publications that are derived from public funding. Publishers have valuable expertise that can help in this process, but only if publishing continues to be sustainable. Such collaboration between publishers and the government is already happening in ways that will increase public access to reports, data, and publications derived from federally funded research. These collaborative initiatives create efficiencies and cost savings for the funding agencies.

Conclusion

AIP believes that a uniform access policy or mandate for scholarly publications would be an ineffective approach. An overarching government-wide policy that would simply mandate a short publication embargo period would fail to take into account such key factors as the specific needs of any given agency, the rapidly changing marketplace and nature of scholarly publishing, and the unique considerations of the various fields of science and the journals that serve them.

The creative and thoughtful discussions that have been spurred by the existing America COMPETES law, organic market forces, and collaborative efforts already underway between publishers and several federal agencies, offer a pragmatic and productive route to success in broadening public access to the all the products of federally funded research: grantee reports, associated data, and the resulting peer-reviewed publications. These efforts reinforce the view of many in the scholarly publishing community that new legislation is not needed at this time.

I believe—and the evidence from the post-COMPETES partnerships shows—that we are making real progress on the interrelated issues of access and interoperability among public and private information platforms and databases.

22 December 2011

Submission for the Record: **Response to November 4, 2011, Federal Register Notice of Request for Information, OFFICE OF SCIENCE AND TECHNOLOGY POLICY, Public Access to Peer-Reviewed Scholarly Publications Resulting From Federally Funded Research; FR Doc No: 2011-28623**

Submitted by: H. Frederick Dylla, Executive Director and CEO, American Institute of Physics
Tel. +1 301-209-3131; Dylla@aip.org

Electronically submitted to: publicaccess@ostp.gov

The American Institute of Physics (AIP) appreciates this opportunity to submit comments and would be delighted to continue working with the Office of Science and Technology Policy (OSTP) and other federal partners through a process of active engagement.

About AIP

The American Institute of Physics is a 501(c)(3) not-for-profit membership corporation created in 1931 for the purpose of “the advancement and diffusion of knowledge of the science of physics and its applications to human welfare.” AIP is an organization of 10 physical sciences societies representing more than 135,000 scientists, engineers, and educators and is one of the largest publishers of scientific information in physics, with activities extending well beyond publishing. AIP delivers valuable resources and expertise in education and student services, science communication, government relations, career services for science and engineering professionals, statistical research, industrial outreach, and the history of physics and other sciences.

As a publisher, AIP plays a central role in the process by which scientific research is developed, communicated, disseminated, and ultimately accepted by the scientific community. AIP publications include 15 journals (three of which are published in partnership with other societies), magazines, including its flagship publication *Physics Today*, and the *AIP Conference Proceedings* series. In addition to its own publication, AIP provides publishing services and expertise to five of its ten Member Societies. To accomplish this, AIP invests millions of dollars annually on peer review, editorial management, production, printing, shipping, distributing, and hosting its archival journals on a fully digital, highly reliable online platform, making the content available at all times to customers around the world in more than 70 countries.

Whether an article is read online or in print, high-quality peer review, page composition (XML), copyediting, and the listing and linking of bibliographic and reference data must be managed, necessitating considerable human capital investment in staff at our Melville, New York, publishing center, in addition to more than 340 editors around the world. Our editors maintain the quality and reputation of our journals, utilizing the well-established system of peer review, whereby independent

experts review submitted articles. Accepted articles are those that pass muster based on established criteria, including novelty and the substantial nature of the research findings. Managing peer review for approximately 30,000 articles submitted to AIP journals every year is a complex undertaking. It requires a large amount of sophisticated electronic resources, associated support personnel, a staff of professional editors—nearly all PhD physicists—and help from tens of thousands of referees. Each year AIP makes such necessary investments to fulfill its public nonprofit mission, generating an intellectual return through the dissemination of scientific research.

Introduction

AIP's highest goal is to achieve the widest possible dissemination of the research results it publishes, including any pertinent associated data and context information. Enabled by Internet technologies, AIP disseminates more information, more widely and more affordably, than ever before in its history, reaching more authors, subscribers, and users than ever before. This accomplishment requires heavy investments in technology and infrastructure (such as an online platform) and business-model innovation to deliver the option of free or low-cost access: open access, pay-per-view, or article rental, recognizing that the value of the final published article needs to be paid for to remain sustainable.

AIP believes that it would be in the best interest of the United States and its government, as well as in the best interest of all other stakeholders, to strike a balance between public access and sustenance of the scholarly publishing industry because of the impact and value it brings to the progress of science and its contributions to American society and economy. Such a balance can be achieved based on shared principles such as the importance of peer review, the recognition of economic realities through adaptable and viable publishing business models, the need to ensure secure archiving and preservation of scholarly information, and the desirability of broad access. One way to achieve this balance is for government to adopt a sensible, flexible, and cautious approach to drafting public access policies—an approach that engages all concerned parties, including federal agencies, scientists, university administrators, librarians, publishers, and the public.

Consistent with the recognition of economic realities, it is AIP's position that government agencies should develop their public access policies through voluntary collaborations with nongovernmental stakeholders, including researchers and publishers. Any policies should be guided by the need to foster interoperability of information across multiple databases and platforms. Agencies' efforts then could be directed toward facilitating cyberinfrastructure and collaboration programs with and between agencies and the stakeholders to develop robust standards for the structure of full text and metadata, navigation tools, and other applications to achieve interoperability across the scholarly literature. More detail on this is provided later in the document. AIP believes that any scholarly publication access policy needs to be flexible to accommodate agency-specific needs and have the capacity to evolve in response to the rapidly changing nature of scholarly publishing.

AIP Responses to RFI Questions

(1) Are there steps that agencies could take to grow the 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 US economic growth and improve the productivity of the American scientific enterprise?

According to trade association and other industry surveys of US publishers, both the nonprofit and commercial sectors serve a robust, innovative global market for the access and consumption of peer-reviewed publications. Academic, corporate, and governmental research and education communities constitute primary segments of the market. Global revenue from scholarly journal publishing was estimated at \$8.0 billion in 2008,^{1,2} with approximately \$3 billion attributed to the US market. The enterprise employs approximately 110,000 people worldwide, with 30,000 in the United States. New publishers, journals, and business models either evolve or emerge constantly, signaling a healthy, competitive marketplace.

The combination of investments in digital and online technology (by publishers as well as others) and the formation of library consortia (assisted by publishers in many cases) in the United States and around the world has accelerated and broadened access to peer-reviewed literature and dramatically decreased cost of such access. AIP serves approximately 2,000 research institutions, and every person affiliated with these institutions has instant access to AIP journal content.

There is a growing presence and diversity of business models in the scholarly market. It is our belief that the government should support and encourage this diversity through its actions and policies through sustainable partnerships with publishers that would contribute to the US economy and maximize the productivity of the scientific enterprise. (For AIP's suggestions of partnerships and pilot projects that would meet mutually beneficial goals and conserve precious federal research funds for the agencies' primary mission of funding research, please see the responses to Questions 4 and 5. These recommendations for partnerships and pilot projects with federal agencies were developed in collaboration with a number of scientific publishers as we engaged over the last year in productive discussions with subject matter experts within the NSF and DOE, two US federal agencies that fund substantial research in the physical and biological sciences and engineering.)

As stated in the 2010 *Scholarly Publishing Roundtable* report,³ many publishers have made the decision to move toward increasingly open structures and archives,⁴ as enabled by open access business models

¹ Cox, J. and L. Cox, *Scholarly Publishing Practice: Academic Journals Publisher's Policies and Practices in Online Publishing*, 3rd ed., ALPSP (2008),

http://www.alpsp.org/ngen_public/article.asp?id=200&did=47&aid=24781&st=&oaid=-1.

² Outsell, "An Open Access Primer-Market Size and Trends" (2009),
http://www.outsellinc.com/contact_us/open_access_primer_2009.

³ Report and Recommendations of the Scholarly Publishing Roundtable, January 2010, available at www.aau.edu/WorkArea/showcontent.aspx?id=10044. Referred to throughout this document as the *Roundtable Report*.

⁴ Morris, S., *Journal Authors' Rights: Perception and Reality* (London: Publishing Research Consortium, 2009),
<http://www.publishingresearch.net/documents/JournalAuthorsRights.pdf>.

and new solutions to associated permissions such as Creative Commons⁵ licenses. These licenses provide a means for exercising certain rights regarding the re-use of an item. For example, these licenses could provide re-use rights if the resulting new works are also made available to the public. The *Roundtable Report* also notes that the number of journals making a change in business model is appreciable but small within the universe of more than 25,000 scholarly peer-reviewed journals.⁶ AIP echoes the *Roundtable Report* assertion that no existing digital business model has demonstrated its viability to the satisfaction of all, and cautions against government endorsement of any single approach.

As part of the market's evolution and scholarly publishers' commitment to community and distribution of results, an increasing number of all types of journal publishers are electing to make their articles freely available to academics and others in 100 or more developing countries. Some well-known programs include the United Nation's HINARI, AGORA, and OARE Research4Life programs, HighWire's Developing Economies Program, and JSTOR's Developing Nations Initiative. Additional programs include those of EIFL, INASP, and TEEAL. For descriptions of these and more, see www.library.yale.edu/~llicense/develop.shtml.

To meet the market's increasing demand for easily accessible, quality information, AIP invests considerably in new technologies for viewing and sharing its journals. Within just the past two years, AIP developed a mobile phone reader for journals, a professional (and freely available) social networking site for physical scientists (www.aipuniphy.org), and an electronic book platform. AIP also launched a multimedia journal on renewable energy (<http://jrse.aip.org>) and one of the first community-style journals in the physical sciences (<http://aipadvances.aip.org>).

Such ongoing investments in existing products and services and the development costs for new products are funded through subscription fees or author payments. AIP and most other scholarly publishers offer an open access option for authors, no matter what type of journal they decided to publish in. Through AIP's Author Select, authors have the option to choose open access for their published article. Less than one percent of authors choose to do so. *AIP Advances*, AIP's newest journal, is an initiative based on community-style review, rapid publication, is fully open access, and employs a Creative Commons license.

This ability for scientific publishers to experiment with different publication, business, and access models is paramount and assures the vitality, diversity, and effectiveness of scholarly communication, leading to scientific and technological advances. Rather than mandate business models and de-incentivize market efficiencies, a more effective approach by government would be to incentivize the continued growth and vitality of the scholarly communication market for the benefit of the scholarly community. To that end, working with publishers, libraries, and other stakeholder communities, research agencies should identify specific needs of particular user groups and collaborate with publishers to meet those needs most effectively. Obviously, researchers, professionals, funders, and various segments of the general public (e.g., patients) have different information needs. AIP is collaborating with other scholarly publishers to identify and address any existing access gaps through

⁵ Creative Commons (<http://creativecommons.org/about>) is a nonprofit corporation that provides free licenses and other legal tools to mark creative work with the freedom the creator wants it to carry, so others can share, remix, use commercially, or any combination thereof.

⁶ Ware, Mark and Michael Mabe, *The STM Report: An Overview of Scientific and Scholarly Journals Publishing*. September 2009.

initiatives such as the low-cost article rental scheme pioneered by DeepDyve, the Research4Life consortium for developing countries (mentioned above), the patientINFORM portal for patients or their caregivers, the Emergency Access Initiative offered to communities affected by natural disasters, and free or substantially discounted access for public libraries, journalists, and high schools.

Nevertheless, based on our experimentation with a modest-cost article rental model (through DeepDyve), AIP remains unconvinced that there is a large unmet demand for public access: only a few thousand members of the general public attempted to access our scholarly content over a year's time, compared to the nearly eight million visitors to AIP content on our online platform.

To maximize the effectiveness of its efforts, government has an important convener role to play in developing standards for data and metadata, and making research more readily searchable and discoverable. Publishers are already working in partnership to develop standardized information and collections through initiatives such as CrossRef.⁷ (For more detail on this, please see response to Question 5.)

With a relatively straightforward implementation of existing policy, government could make the funder-collected and maintained outputs of taxpayer-funded research, such as grant reports and research progress reports, freely available to the public.⁸ Furthermore, to incentivize open access publishing, funds could be made available specifically to support payment for open access to published articles as pilot projects. Several research funders already do this (Howard Hughes Medical Institute, Wellcome Trust, and Max-Planck Institutes).

In the same vein, government funding could be provided to license content from publishers in order to make it available to specific audiences. (Publishers license content to customers of many kinds, including government agencies, and have the ability to ensure its continued availability with existing infrastructure.)

AIP has been a leading participant in organizing working groups that are proposing and planning partnerships with NSF and DOE on access, linking of grantee reports to publications, data mining across agency-publisher databases, tools and methodology for identifying publicly funded work, and potential pilot projects in the above areas. (More detail on this can be found in response to Question 5.)

Government mandates for public access come at a significant cost to the US economy and to the scientific enterprise. The National Institute of Health's (NIH) PubMed Central (PMC) data indicates that two-thirds of its users are from overseas. This suggests that critical export opportunities for the industry may be compromised, resulting in loss of US jobs. Significant economic value added generated by the publishing industry could be wasted if revenue derived from sales in the global market is compromised

⁷ CrossRef (www.crossref.org) is a not-for-profit group founded by publishers in 2002 and maintains 50 million items. Almost 1000 publishers participate, assigning Digital Object Identifiers (DOIs) to published content items. Development of the CrossRef service has resulted in seamless navigation of the research literature by users so that researchers using the bibliography in one article can link from a reference to the full text of the referenced article.

⁸ This would ensure readability to the broadest audience. NSF is already pursuing such a policy, see <http://www.nsf.gov/pubs/policydocs/pofaqs.jsp>, and DOE through its Office of Scientific and Technical Information provides public access to nearly 300,000 DOE-funded research reports, see <http://www.osti.gov/bridge/>.

or eliminated. Furthermore, mandates often result in additional costs for publishers. For example, although only a very small portion of AIP's content is subject to the NIH public access mandate (AIP is primarily a physical science publisher), AIP had to incur costs to modify formats and procedures in order to deposit manuscripts into PMC. AIP remains concerned that PMC is shifting readers from the publishers' sites to PMC despite linking arrangements, thus undermining the value of the publishers' investments.

AIP has concerns about any government policy affecting global trade balance. The number of papers submitted to AIP journals from China exceeded the US submissions two years ago. In response, AIP opened an editorial and marketing office in China to help promote established AIP journals in the physical sciences in China, rather than see China develop competing international journals. Free and unimpeded access to US journal content, even if one were to factor in a short embargo period, will undermine our and other US publishers' needed revenue to establish business relationships in potentially lucrative and large global markets such as China.

In summary, AIP believes that publishers should continue to be free to experiment with various business models in the marketplace of ideas and economics. AIP endorses the *Scholarly Publishing Roundtable* recommendation that "Agency policies should encourage the development, in a competitive landscape, of new value-added information products and services that take advantage of a scholarly environment in which articles are increasingly interoperable and available through licenses that support creative reuse. Such development should be carried out on a level playing field among all those who would devise such products and services." We believe that it is essential that any public access process does not undermine the ability of the market to create and sustain peer-reviewed journals.

(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?

Scientific publishers, such as AIP, rely heavily on the reputation of their journals to compete in the marketplace. Copyright protection reinforces the motivation for sustaining managed peer review, thereby protecting a journal's reputation. Any policy decisions regarding the publication and dissemination of peer-reviewed scholarly publications resulting from federally funded scientific research must respect US copyright law as it presently exists. Under the law, these works meet the criteria for copyright protection. It is a constitutional right granted to the copyright holder to exercise the exclusive rights attached to a work. In its role as the guardian of those rights, government must seek to strike the appropriate balance for all stakeholders through fair interpretation of the law.

It is AIP's position that agencies should provide free public access to final research reports and link to the peer-reviewed journal articles, which are available through a variety of access mechanisms. This solution would drive the standardization of information reported on publicly funded research, promote rapid dissemination (rather than waiting for an article to be authored and subsequently peer reviewed), and ensure preservation of intellectual property rights, which provide the incentive for producing, distributing, and preserving all forms of intellectual property.

AIP encourages agency policies and actions that work to ensure copyrighted materials are protected from unauthorized dissemination and piracy. Copyright is an essential ingredient in promoting creativity, innovation, and the continued integrity and reliability of the scholarly record. There is some evidence that the NIH policy undermines intellectual property rights and promotes piracy of intellectual property. As noted in response to Question 1, the NIH public access policy and availability of articles through NIH's database, PMC, undermine an important US export market. Furthermore, copyrighted material downloaded from PMC appears on rogue Internet sites, resulting in millions of dollars in annual losses to US publishers.

Nearly all scholarly publishers adopt liberal copyright policy, allowing authors to post copies of their manuscript on their individual and institutional websites with very little restriction, share copies with colleagues, and to use their manuscripts for other educational and research purposes. Only commercial use is restricted and enforced by the industry.

(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?

Although a centralized data platform has some obvious advantages of simplicity of operation, the use of a centralized, government-controlled platform for a large corpus of scholarly content has significant downsides, including increased costs to the government. A centralized approach discourages innovation by driving traffic away from innovators, including publishers, thus minimizing scientific and commercial opportunities. However, an important role for government in this arena would be to drive and fund the development of interoperability standards and promote the widespread use of such standards.

AIP supports the recommendation of the *Roundtable Report* that states that government policies should be guided by the need to foster interoperability and encourages “. . . additional multiagency programs supporting research and development to expand interoperability capacity and to develop and promote additional interoperability practices and standards.” The *Roundtable Report* further notes that the National Science Foundation, the Department of Energy, and other agencies provide important funding for the development of interoperability capacities through their cyberinfrastructure programs.

In developing public access policies and procedures, agencies should carefully consider international cooperation with a larger vision that includes building standards and fostering distributed systems that are global in scope and go far beyond the work funded by US federal research dollars. In the Internet age, research and research resources are distributed globally. US federally funded research is only one part of the entire universe of information on any given topic, and in some disciplines, research is increasingly non-US government funded. A centralized repository such as PubMed Central, though by some measures successful, is not a model that is universally applicable or necessarily the best model for the future. Indeed, the success of the World Wide Web is its evolving capability to connect an exponentially growing array of highly distributed information resources and databases. Any successful

and optimized scientific publishing system will incorporate effective incentives to implement and expand interoperability and reuse across internationally distributed databases.

It is AIP's position that stewardship of publications in the Internet age should be the collaborative responsibility of the publishing, library, and research communities. US government involvement in the long-term stewardship of publications is best addressed as part of the copyright system and through the Library of Congress digital preservation initiatives primarily as a promoter of standards, as noted above, and as one of many stewards of specific data platforms that need to be linked across public and private boundaries.

What constitutes a publication and the nature of publication is changing with technology. A publication is no longer just a chunk of text fixed in time forever but a fluid representation. Publications can include supplemental material, multimedia files, software, links to resources on the web, and can be revised and corrected over time by the authors and publishers, hence the emergence of new community initiatives such as CrossRef's CrossMark⁹ service, which electronically watermarks an article's Version of Record (VoR), and DataCite,¹⁰ which extends the CrossRef-promoted Digital Object Identifier (DOI) to datasets. Any plan for the future should recognize that the static aggregation/library model is not likely to hold up well in the distributed and dynamic Internet milieu.

AIP believes that it is unlikely that one optimal procedure for preservation and stewardship would emerge to become applicable across all of scholarly publishing. For now, AIP strongly recommends that agency policies embrace diversity, decentralization, and interoperability. In the long term, systematic collaborations among stakeholders (government, publishers, universities and their libraries, and other not-for-profit participants in the scholarly publishing system) will be necessary to achieve maximum benefit. We note that libraries, in partnership with publishers, have established entities for preservation of digital documents that are already in wide use, for example, Portico¹¹ and CLOCKSS.¹²

Long-term stewardship of content comes at significant cost that is being borne by publishers. In an era of dwindling federal resources, central federal repositories are duplicative, an unnecessary expense, and a recurring burden that may not be viable in the short or long term. Long-term stewardship might be more suitably carried out by the private sector or through collaborative stakeholder projects. There are productive ways to define appropriate roles of government and nongovernmental participants in the system, and ways that government agencies and nongovernmental stakeholders can collaborate as

⁹ CrossMark (www.crossmark.com) is a current pilot project of CrossRef to that will allow readers to easily determine whether they are looking at the publisher-maintained, stewarded version of a journal article.

¹⁰ DataCite (<http://datacite.org>) is a not-for-profit organization established to facilitate easier access to research data on the Internet, increase acceptance of research data as legitimate, citable contributions to the scholarly record, and support data archiving that will permit results to be verified and re-purposed for future study.

¹¹ Portico (<http://www.portico.org/digital-preservation/>) is a digital preservation service provided by a not-for-profit organization with a mission to help the academic community use digital technologies to preserve the scholarly record and to advance research and teaching in sustainable ways. It is among the largest community-supported digital archives in the world, working with libraries, publishers, and funders to preserve e-journals, e-books, and other electronic scholarly content.

¹² CLOCKSS (*Controlled* LOCKSS) is a not-for-profit joint venture between the world's leading scholarly publishers and research libraries whose mission is to build a sustainable, geographically distributed dark archive with which to ensure the long-term survival of web-based scholarly publications for the benefit of the greater global research community (<http://www.clockss.org/clockss/Home>).

equal partners to their mutual benefit in strengthening the scholarly publishing system and expanding public access to its outputs.

(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?

Yes, please see detailed response to Question 5 below.

(5) What steps can be taken by federal agencies, publishers, and/or scholarly and professional societies to encourage interoperable search, discovery, and analysis capacity across disciplines and archives? What are the minimum core metadata for scholarly publications that must be made available to the public to allow such capabilities? How should federal agencies make certain that such minimum core metadata associated with peer-reviewed publications resulting from federally funded scientific research are publicly available to ensure that these publications can be easily found and linked to federal science funding?

To facilitate public access and drive and support scholarship, agency databases should be able to communicate with each other. Each agency's policies should include common core properties that promote access to and interoperability among the content in all public access databases. Specifically, AIP encourages agencies to develop collaborations and partnerships with scientific publishers to develop and implement:

- Standards and persistent identifiers to enhance the discoverability of research results and to promote interoperability among agency, publisher, and any third party databases and platforms;
- Discovery tools to facilitate journal content mining; and
- Pilot projects that would drive access, use, and innovation from research results.

Specifics on these items are discussed below.

Beyond common properties, agencies should have the flexibility to manage and modify their policies in response to evolving circumstances. Agencies should fully engage researchers, institutions, and publishers working in fields that coincide with the agencies' missions, both in establishing initial public access policies and in modifying those policies as appropriate over time.

Many scholarly publishing organizations, such as AIP, were founded by scientists for scientists and fully embrace providing publishing and other services as their primary mission. As part of this, AIP's CEO was an active member the Scholarly Publishing Roundtable and subsequently helped organize working groups of nonprofit and commercial publishers to propose and implement joint projects with both the DOE and NSF with a mutually agreed-upon goals. Moreover, AIP is a cofounder of CrossRef and participates in a number of standards organizations such as the National Information Standards Organization (NISO—www.niso.org), National Federation of Advanced Information Services (NFAIS—www.nfais.org), and the newly formed consortium Open Researcher and Contributor ID (ORCID—www.orcid.org), with a purpose to develop unique researcher identifiers.

Standards and Identifiers: Agency Funding Information

Most funding agencies currently require researchers to acknowledge in publications the support that they have received. There are no standards, however, on how this should be done. Consequently, agency funders find it difficult to know what publications have arisen from the research they have funded. AIP has promoted the recommendation that publishers develop, in collaboration with funding agencies and CrossRef, a means of standardizing funder information and make that information available to funding agencies and the public. We believe that a community-wide solution of this type will be easier and far less expensive to deliver than for each agency to develop its own response to the problem. This is because publishers are in the best position to provide a simple way of ensuring that journal articles are accompanied by standardized, high-quality metadata providing information about the agency, program, and the specific grant that funded the research. It would be very expensive for agencies to obtain this information through data mining existing publisher databases.

This proposal has been endorsed by CrossRef and the major scientific, technical, and medical (STM) publishing trade associations: the Professional and Scholarly Publications Division of the American Association of Publishers (PSP-AAP) and the International Association of Scientific Technical and Medical Publishers. Related to this proposal, the DOE's Office of Scientific and Technical Information (OSTI) has agreed to maintain a registry of standard nomenclature for funding agencies and the associated naming and numbering system for grants. OSTI already houses technical reports and data sets for more than 40 federal and international funding organizations.

With the successful implementation of this funding identity proposal by STM publishers and CrossRef, agencies would have access to standard metadata from published articles. By displaying this information on agency websites, visitors—from the research community to the general public—could follow the link [enabled through the Digital Object Identifier (DOI)] to the publisher's platform where article abstracts are freely available and the Version of Record (VoR) (maintained by the publishers) is available through a variety of access mechanisms, including innovative rental access models, which give the public instant access for a modest fee. More than 40 scholarly publishers are currently testing this access mechanism.

Standards and Identifiers: DOIs for Data Sets and Supplementary Material

Increasingly throughout the world, grant investigators are being asked to share or provide plans regarding how they will share with other researchers the primary data, samples, physical collections, and other supporting materials created or gathered in the course of their work. Grantees are expected to encourage and facilitate such sharing. Scholarly publishers are already participating in a number of initiatives designed to facilitate the voluntary sharing of data or to foster interoperability among data sharing repositories, and they would be willing to work with NSF, DOE, and other database/repository operators to develop recommended practices for assigning DOIs to data sets and supplementary material.

For data policies, publishers would draw on their experience with initiatives such as Opportunities for Data Exchange (ODE; see www.alliancepermanentaccess.org/current-projects/ode), which aims to gather and promote best practices on the way scientific data are treated, and CoData, a partner of the International Council for Science (ICSU) World Data System (www.icsu-wds.org). The goals of the relatively new ICSU World Data System (WDS) are to create a global federated system of long-term data archives and data-related services covering a wide spectrum of natural sciences, thereby encouraging

interdisciplinary scientific approaches. For supporting information, publishers would draw on their involvement with the joint NISO/NFAIS Working Group on Supplementary Journal Information (see www.niso.org).

Standards and Identifiers: Author Disambiguation

Name ambiguity and attribution are persistent, critical problems embedded in the scholarly research ecosystem. AIP encourages agencies to work in collaboration with publishers as well as universities, funding organizations, and corporations from around the world to eliminate this problem through ORCID. ORCID is a newly established nonprofit organization whose goal is to establish an open, independent registry of researchers that is adopted and embraced as an industry-wide standard to resolve systemic name ambiguity by means of assigning unique identifiers linkable to an individual's research contributions. Researchers will be able to create, edit, and maintain an ORCID ID and profile free of charge and will define and control the privacy settings of their own ORCID profile data. Participants expect that accurate identification of researchers and their work will facilitate emergence of new services and benefits for the research community by all types of stakeholders in scholarly communication: from commercial actors to nonprofit organizations, and from governments to universities.

Such a standard will not only enhance the scientific discovery process but also improve the efficiency of funding and collaboration. Participation in ORCID is open to any organization that has an interest in scholarly communications. All profile data contributed to ORCID by researchers or claimed by them will be available in standard formats for free download (subject to the researchers' own privacy settings) that is updated once a year and released under a Creative Commons license. All software developed by ORCID will be publicly released under an open-source software license approved by the Open Source Initiative (OSI). For the software it adopts, ORCID will prefer open source. ORCID is governed by representatives from a broad cross section of stakeholders, including publishers, library organizations, research institutions, and funding agencies (see <http://orcid.org/board-of-directors>).

Discovery Tools: Content Mining

Content mining can be especially useful to the scientific community in driving interdisciplinary research and supporting the identification of new areas of discovery, and publishers are committed to managing content in modern digital formats to ensure that users gain maximum benefit. Scholarly publishers should work with funding agencies to develop pilot projects for journal content mining that would create thesauri, using their expertise to identify, organize, and analyze content to create conceptual links within and between highly technical subject matter. Although there are various ways to perform this type of processing, certain elements are common to all methods, including an automated way to process all sizes and types of content in which to identify relevant information and facilitate its extraction and analysis.

Such pilots should focus on goals such as the following:

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

Consensus approaches within the community could also be explored for developing better standardized, mining-friendly content formats, a shared content mining platform, and common permission rules for content mining. The Publishers Research Consortium recently completed a study on article-level content mining based on a broad survey of ongoing or planned activities among nearly 30 STM publishers or associations (see www.publishingresearch.net/documents/PRCSmitJAMreport20June2011VersionofRecord.pdf).

Pilot Projects: Sponsored Access to Published Research

The “Gold Open Access” dissemination model, which includes an article processing charge paid by the author or their institution, delivers immediate and unrestricted online access to the final published article (defined by NISO as the Version of Record).

AIP suggests that agencies could work with publishers to set up experiments to answer the following questions dealing with the cost, benefits, and sustainability of the Gold Open Access model, as well as investigate how such a model should be funded and administered:

- How much would it cost an agency to fund Gold Open Access in the aggregate and on a per-article basis?
- What is the most effective method to provide Gold Open Access funding for authors? The ability to use grant funds for sponsorship? A separate pool of funding reserved solely for Gold Open Access sponsorship? Other means?
- Should authors be required to expend grant funds on publishing of their articles? If not, how can authors be encouraged to utilize the available funds? (Several methods/messages could be tested.)
- How can agencies best administer a Gold Open Access program?
- Does Gold Open Access offer agencies new opportunities to showcase the productivity of their funding activities to the American public and federal oversight committees?

Pilot Projects: Linking to/from Research Reports

AIP encourages federal agencies to fund a pilot project that would seek to determine whether and how publisher content derived from agency-funded research could be mapped against agency research reports and other content. Specifically, the pilot would send users from publisher websites to the agency website to view free government-sponsored research reports and would, likewise, send users from the agency websites to publisher sites to view free abstracts and links to the Version of Record of articles connected to a particular research report or funded project.

If successful, this would result in interoperability between onsite agency content and publisher platforms. This is of interest to scholarly publishers because they would like to work with major research funders to identify, organize, evaluate, and highlight published results from federally funded research, as well as identify relationships, projects, and offerings that might be applicable to other research funders.

Possible outcomes of the pilot could include:

- The ability to identify all agency-funded research within publisher offerings and the ability to deliver associated metadata to agencies,
- The ability to establish mechanisms and approaches that could be implemented (for all research funders) across the industry,

- A capability to report to major funders on the impact of the research they fund, e.g., through bibliometric and other tools,
- A “research dashboard” capability or the ability to contribute to one already in existence, e.g., <http://rd-dashboard.nitrd.gov/>,
- A mechanism for low-cost content rental access to published articles (Versions of Record) and a mechanism to explore its impact,
- Subject area content portfolios of agency-funded research articles for internal agency use (e.g., study sections),
- The possibility to use the DOE-OSTI platform (the <http://www.science.gov/>) to extend this pilot to other federal funding agencies, and
- Models to illustrate how traditional publishing systems can coexist with self-archiving.

(6) How can federal agencies that fund science maximize the benefit of public access policies to US 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?

An excellent mechanism to ensure public access to federally funded research results is by providing access to final agency reports. Every federally funded research project is required by law to provide a detailed final report. The research reports are a condition of the government contract. These reports should be archived and made accessible to the public. Some science funding agencies make these reports freely available via the web, others do not. Making all such reports available and accessible in a comprehensive and systematic way would solve an essential public access problem. One leading example is DOE’s Office of Scientific and Technological Information, which publishes final reports online in a portal called Information Bridge. These reports are not journal articles, but the final reports are often much longer than the resulting journal article (if such article exists—researchers typically publish only positive results and then have to meet the publication standards of the journals in their field) and provide more information.

Moreover, NSF instituted a new reporting requirement as a result of specific legislation in the America COMPETES Act (Section 7010: Reporting of Research Results), which required 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 Website.” For several years, publishers have proposed working with authors to develop short abstracts for a lay audience to accompany each research report.

Publishers are partnering with federal agencies to develop policies that maximize public access to research results and provide easy links between research reports (detailing research results, perhaps including lay summaries) and the peer-reviewed Version of Record, including complete access to the abstract or summary. Such projects would result in interoperability between funder and publisher content, ensuring access and better reporting on the results of funding.

In addition, please see the response to Question 5 above for specific agency initiatives.

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

No. Publishers also invest in these other types of content used by researchers, often by conceptualizing the project, commissioning the content, and investing heavily in its development. Any kind of mandated access to that content is an expropriation of that content.

(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?

AIP believes that a uniform access policy or mandate for scholarly publications would be an ineffective approach. An overarching government-wide policy or embargo period would fail to accommodate such key factors as the specific needs of any given agency, the rapidly changing nature of scholarly publishing, and the unique considerations of the various fields of science and the journals that serve them.

AIP analyzed related industry data using the “cited half-life” metric as a relative indicator for how long journal titles within scientific categories are being accessed and cited, thus reflecting economic viability. The findings could help inform considerations related to embargo periods. Based on the evidence related to AIP journals and to journals covering physics and related sciences, significant economic uncertainty remains with the assignment of minimum embargo periods. In looking at a sample of several physics and related topics and AIP journals within those categories, AIP found that physics journals have a longer cited half-life compared to some other scientific disciplines, and furthermore, AIP Journals have a longer cited half-life than their respective physics category averages.

The chart on the following page provides are some examples.

Cited Half-Life of AIP Journals/Physics vs. Other Scientific Disciplines

The cited half-life for the journal is the median age of its items cited in the current year. Half of the citations to the journal are to items published within the cited half-life.

| Scientific Categories | Average Cited Half-Life within Sub-Category (Years) | Average Cited Half-Life of AIP Journals within Sub-Category (Years) | AIP Journals in Category |
|--|---|---|--|
| Applied Physics | 5.6 | 7.8 | <i>Applied Physics Letters, Journal of Applied Physics, Journal of Low Temperature Physics, Review of Scientific Instruments</i> |
| Chemical Physics | 7.1 | >10 | <i>The Journal of Chemical Physics</i> |
| Physics—Fluids & Plasmas | 6.6 | 7.6 | <i>Physics of Fluids, Physics of Plasmas</i> |
| Mathematical Physics | 6.5 | >10 | <i>Journal of Mathematical Physics</i> |
| Medicine, Research and Experimentation | 5.4 | | |
| Cardiac & Cardio Systems | 4.9 | | |
| Emergency Medicine | 5.7 | | |
| Robotics | 5.5 | | |

Source: Thomson Reuters, ISI Web of Knowledge, Journal Citation Reports, Year 2010

In lieu of trying to solve the public access problem by imposing a one-size-fits-all solution with a fixed embargo length for all articles that have some component of federal funding or introducing a complicated scheme for varying embargo lengths (as necessary to address field-specific conditions), AIP proposes a simpler system that allows government to accomplish public access in a way that is not only effective, efficient, and sustainable, but also keeps the US scientific enterprise thriving as it moves into the future.

To summarize the key components, AIP and a number of our colleagues from the scientific publishing community propose the following scheme to improve public access to the results of publically funded research:

1. Scholarly publishers as a group have proposed modifications to their author submission software so that all journal articles written after the implementation date would include funding agency information along with the standard metadata that is already being deposited in CrossRef and other standard bibliographic databases. This new metadata, which specifically tags the funding agency(s) responsible for the research leading to the journal article, would be deposited in the CrossRef database. (The CrossRef database has been developed and

maintained by this nonprofit consortium for the past 12 years and now contains the metadata for more than 50 million scholarly articles and related content.) Funding agencies can procure a license to this database at modest cost—many already have. Such a license provides access to the article metadata, including the critical article identifier (the DOI).

2. With the successful implementation of this funding identity proposal by scholarly publishers and CrossRef, agencies would have access to the standard metadata from published articles. By displaying this information on agency websites, visitors—from the research community to the general public—could follow the link [enabled through the Digital Object Identifier (DOI)] to the publisher’s platform where article abstracts are freely available and the Version of Record (VoR) (maintained by the publishers) is available through a variety of access mechanisms, including innovative rental access models, which give the public instant access for a modest fee. More than 40 scholarly publishers are currently testing this access mechanism.
3. Scholarly publishers have proposed and initiated pilot projects with funding agencies to link agency research reports and related content on agency sites to publisher content tagged with the same funding information, thus expanding interoperability between agency and publisher databases and access to the linked content.

22 December 2011

Submission for the Record: **Response to November 4, 2011 Federal Register Notice of Request for Information, OFFICE OF SCIENCE AND TECHNOLOGY POLICY, Public Access to Digital Data Resulting From Federally Funded Scientific Research; FR Doc. 2011-28621**

Submitted by: H. Frederick Dylla, Executive Director and CEO, American Institute of Physics
Tel. +1 301-209-3131; Dylla@aip.org

Electronically submitted to: digitaldata@ostp.gov

The American Institute of Physics (AIP) appreciates this opportunity to submit comments and would be delighted to continue working with OSTP and other federal partners through a process of active engagement.

About AIP

The American Institute of Physics (AIP) is a 501(c)(3) not-for-profit membership corporation created in 1931 for the purpose of “the advancement and diffusion of knowledge of the science of physics and its applications to human welfare.” AIP is an organization of 10 physical sciences societies representing more than 135,000 scientists, engineers, and educators. As one of the largest publishers of scientific information in physics, AIP employs innovative publishing technologies and offers publishing services for its Member Societies. AIP's suite of publications includes 15 journals, three of which are published in partnership with other organizations; magazines, including its flagship publication *Physics Today*; and the AIP Conference Proceedings series. AIP delivers valuable resources and expertise in education and student services, science communication, government relations, career services for science and engineering professionals, statistical research, industrial outreach, and the history of physics and other sciences.

Enabled by Internet technologies, AIP disseminates more information, more widely and more affordably, than ever before in its history, reaching more authors, subscribers, and users than ever before. This accomplishment requires heavy investments in technology and infrastructure (such as an online platform) and business-model innovation to deliver the option of free or low-cost access: open access, pay-per-view, or article rental, recognizing that the value of the final published article needs to be paid for to remain sustainable.

Introduction

AIP's highest goal is to achieve the widest possible dissemination of the research results it publishes, including any pertinent associated data and context information. As a scholarly publisher, AIP believes that better discoverability and reuse of original research data are to be encouraged at all levels and among all stakeholders. AIP also believes that data resulting directly from federally funded scientific

research should be made freely available in a sustainable manner and that this is best achieved through appropriate policies that leverage public-private collaboration.

AIP believes that it would be in the best interest of the United States and its government, as well as in the best interest of all other stakeholders, to strike a balance between public access and sustenance of the scholarly publishing industry because of the impact and value it brings to the progress of science and its contributions to American society and economy. Such a balance can be achieved based on shared principles such as the importance of peer review, the recognition of economic realities through adaptable and viable publishing business models, the need to ensure secure archiving and preservation of scholarly information, and the desirability of broad access. Policies should recognize that hosting, maintaining and preserving raw data or data sets, and continuing to make such data available over the long term, has a cost, which, in certain circumstances, the host should be entitled to recover. One way to achieve this balance is for government to adopt a sensible, flexible, and cautious approach to drafting public access policies—an approach that engages all concerned parties, including federal agencies, scientists, university administrators, librarians, publishers, and the public.

Consistent with the recognition of economic realities, it is AIP's position that government agencies should develop their public access policies through voluntary collaborations with nongovernmental stakeholders, including researchers and publishers. Any policies should be guided by the need to foster interoperability of information across multiple databases and platforms. Agencies' efforts then could be directed toward facilitating cyberinfrastructure and collaboration programs with and between agencies and the stakeholders to develop robust standards for the structure of full text and metadata, navigation tools, and other applications to achieve interoperability across the scholarly literature. More detail on this is provided later in the document. AIP believes that any scholarly publication access policy needs to be flexible to accommodate agency-specific needs and have the capacity to evolve in response to the rapidly changing nature of scholarly publishing.

AIP specifically recommends that federal grants set aside funds to support researcher data management and deposit efforts. Federal agencies could also play a role in supporting and encouraging the establishment of discipline-specific data archives where these are currently lacking. The amount and type of support should be determined in collaboration with key stakeholders involved in the deposit, storage, and preservation of data.

Federal policies should also focus on supporting and encouraging the development of community standards for the citation and reuse of data sets, thereby facilitating the creation of a system that gives researchers an incentive to share data resulting from federal grants.

AIP Responses to RFI Questions

Preservation, Discoverability, and Access

(1) What specific Federal policies would encourage public access to and the preservation of broadly valuable digital data resulting from federally funded scientific research, to grow the U.S. economy and improve the productivity of the American scientific enterprise?

We would make the distinction that it is not “public access” in the broadest sense that is important but rather access by other scientists who can use the digital data for the further advancement of science.

As data are not copyrightable, policies about access become policies about deposit by the data owner or proxy into an accessible system. It should be noted, though, that any policies should recognize and take into account differences between ‘databases’ (information products created for the specific display and retrieval of data) and ‘data sets’ (sets or collections of raw relevant data captured in the course of research or other efforts). Policies could require that data generated from federally-funded research be deposited in a certified and openly accessible repository; furthermore, researchers could be encouraged to make these deposits upon submission of their first manuscript showing results that were based on the data set. Although some agencies already have a preservation/access role (for example, DOE Order 241.1B), AIP agrees with the Interagency Working Group on Digital Data that “data stewardship is best accomplished in a system that includes distributed collections and repositories maintained where the custodian has trusted community-proxy status with the relevant communities of practice.” Agency policies should support and encourage such a distributed system for both access and preservation; that is, policies should recognize and build upon the broad set of capabilities that exist for both access and preservation within the library and publishing communities for both documents and data, such as Portico, LOCKSS.

The integrity of preserved data would also need to be taken into account and supported by any policy.

(2) What specific steps can be taken to protect the intellectual property interests of publishers, scientists, Federal agencies, and other stakeholders, with respect to any existing or proposed policies for encouraging public access to and preservation of digital data resulting from federally funded scientific research?

All policies should comply with current copyright and patent law. Data should be embargoed to the principal researcher until conclusions drawn from the data can be published in the research literature. An additional maximum embargo of one year would also provide for the filing of patents by the grantees (or their institution) as allowed by many, if not all, funding agencies (HR 1249 Sec 102(b)(1)(A)). See also the distinction between databases and datasets as addressed response to question 1.

(3) How could Federal agencies take into account inherent differences between scientific disciplines and different types of digital data when developing policies on the management of data?

Differences between scientific disciplines and types of digital data must be taken into account by domain experts at the time of proposal review (note the language used in the Data Management Plan FAQ’s of NSF in a variety of instances: “to be determined by the community of interest through the process of peer review and program management.”) Only such experts will be able to determine if the data to be generated by the proposed research will be of longer term value to the scientific community of interest and if its type conforms to acceptable community standards.

Metadata—data about the data—which would include information both about what the data is and how it was collected, is addressed further in this response.

(4) How could agency policies consider differences in the relative costs and benefits of long-term stewardship and dissemination of different types of data resulting from federally funded research?

Policies must first recognize that not all data is worth preserving. Each type of data should be assessed regarding long-term stewardship. Policies would have to take into account not just the size of the datasets but also long-term usability, which depends on the rate of technology change, and level of documentation required. Along with the data, enough information needs to be preserved to reproduce the dataset. As noted in the answer to question 3, agencies will need to call upon data experts as well as scientific experts.

(5) How can stakeholders (e.g., research communities, universities, research institutions, libraries, scientific publishers) best contribute to the implementation of data management plans?

There needs to be an interconnected system for access to and sharing and preservation of data based on community-developed standards and best practices. The system needs to encourage innovation and must support multiple solutions—data as an information resource is inherently more complicated than scholarly articles. Each stakeholder will then need to contribute based on their specific skills and expertise. Libraries, through Institutional Repositories, could take on a stronger preservation role. Publishers have been adding value to the research process and providing access to and preservation of the scholarly literature for hundreds of years and could extend this to data, well beyond current support for supplemental material. Universities and research institutions have both scientific domain knowledge and data and information experts. Any system will need to preserve incentives for innovation.

Consider, for example, work being done by the Data Preservation Alliance for Social Sciences through their partnership with the Library of Congress, LOCKSS, and Dataverse to prototype a policy-based replicated data archive.

Other examples include:

- linking between datasets and their resulting scholarly publications based on community-accepted standards, thus ensuring datasets become part of the scientific literature;
- Having clear standards and guidelines for the certification and auditing of data repositories; encouraging a system that incentivizes data repositories to maintain the accuracy or integrity of the data once it has been deposited;
- Incentivizing the deposit of datasets and ensuring that the administrative burden this imposes on researchers minimal.

(6) How could funding mechanisms be improved to better address the real costs of preserving and making digital data accessible?

Require data management plans and coordinate plan requirements across agencies and to community standards (see the Open Archive Information System Reference Model – ISO standard 14721:2003). What constitutes data that needs to be preserved should be clearly identified through the process of peer review and program management. Preserving and disseminating digital data should then be considered “part of the cost” of funding and doing research, not “an additional cost”. Funding agencies could emphasize that proposals must take into account data fit for reuse and preservation. Again, this

should be the approach across agencies. Research labs/institutions/university overhead rates would need to include costs of data preservation.

As pointed out in the final report from the Blue Ribbon Task Force on Sustainable Digital Preservation and Access (*Sustainable Economics for a Digital Planet*): “Policy mechanisms can play an important role in strengthening weak motivations” as there is often “misalignment of incentives between communities that benefit from preservation (and therefore have an incentive to preserve), and those that are in a position to preserve (because they own or control it) but lack incentives to do so.”

(7) What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?

If data is created in the course of federally-funded research, then the funding agency could require that any such data deemed to be “preservation data” be deposited in a recognized archive. Through direct agency involvement in creating a “comprehensive framework for data access and preservation” based on community-accepted standards and best practices for data citation and reuse, agencies would maintain lists of certified repositories. Certified repositories could be similar to the data center members of the DataCite organization (of which DOE’s Office of Scientific and Technical Information is a member) or participants in the SafeArchive program of Data-PASS. In addition, grantee data management plans could be required to identify all datasets expected to be produced from funded work.

Certification of compliance would then simply require grantee reporting to include in reports on their funded proposal the data citations and the repository where the data was deposited.

As work is already being carried out to develop standards in this area (i.e. *The ISO 16363 Standard for Trusted Digital Repositories*), it would be more expedient for federal agencies to work within and help support such standards.

(8) What additional steps could agencies take to stimulate innovative use of publicly accessible research data in new and existing markets and industries to create jobs and grow the economy?

AIP agrees with the statement from the Interagency Working Group on Digital Data (IWGDD) in its report, *Harnessing the Power of Digital Data for Science and Society*, that “the current landscape lacks a comprehensive framework for reliable digital [data] preservation, access, and interoperability”. We feel that there is a very important role for the federal government and its science funding agencies to play to help create and promulgate such a comprehensive framework.

Federal investment in creating stable, standardized, and accessible data will be an essential base from which innovation can occur. The ease of reuse could then lead to developments akin to IBM Research’s “Many Eyes” product for data visualization (www-958.ibm.com), or spur the private sector to offer data services for researchers.

(9) What mechanisms could be developed to assure that those who produced the data are given appropriate attribution and credit when secondary results are reported?

This ecosystem of attribution and credit already exists with respect to scholarly articles. A researcher's standing in their field is largely a result of their list of authored scholarly articles and the number of citations to those published articles. The credit comes in the form of respect from peers, funding for further work, and career advancement, and rests in large part on the underlying quality control provided by peer review. Not providing appropriate attribution is considered unethical scientific behavior and can lead to the retraction of published work.

The mechanisms to be developed would support an extension of this system to cover data. The elements to support are:

- data must be recognized as a primary research output,
- data must have unique and persistent identifiers and be fully citable, thereby allowing its use and reuse to be tracked and recorded in the same way as scholarly publications, and
- data citation information must be used for research evaluation and reward.

Persistent identifiers for data could be handled through use of digital object identifiers already used for scholarly articles or similar (see Datacite.org). There are also examples of recommended practice for citing data. [For example: creator (publication year): Title, Publisher, identifier; see <http://datacite.org/whycitedata> and DOE's Data ID Service.]

Publishers could support the development of such a system by requiring that all data needed to reproduce the results and conclusions of a published scholarly article must be cited according to community standards.

Funding agencies could support the development of such a system by recognizing data that has been archived and made available to the research community as "first class research objects" at the same level as articles. Agencies should also recognize any reuse of these data which could then be counted via citations.

See the Australian National Data Center's "Building a Culture of Data Citation" poster available at <http://ands.org.au/cite-data/index.html>.

For a hybrid example spanning the world of digital data and scholarly publication, see the *Journal of Physical and Chemical Reference Data*, a long and successful collaboration between AIP and the National Institute of Standards and Technology.

Standards for Interoperability, Reuse and Re-Purposing

(10) What digital data standards would enable interoperability, reuse, and repurposing of digital scientific data? For example, MIAME (minimum information about a microarray experiment; see Brazma et al., 2001, *Nature Genetics* 29, 371) is an example of a community-driven data standards effort.

First, it is important to separate metadata standards from data format standards. Metadata standards could be developed that are lightweight enough to be widely interoperable and extensible so as to accommodate discipline-specific needs (within the XML publishing standard). These standards would need to cover both bibliographic information (data creator, date of creation, what the data describes, where it can be accessed, etc.), and how it was collected (experimental apparatus, experimental conditions, location, etc.).

Data format standards that would enable reuse and repurposing would need to be developed at the discipline-specific level. There need not be one solution per discipline: it may be that the communities in question need a handful of solutions that correspond to the various types of data and/or modes of scientific research that produces the data. So while it is true that actual data solutions need to be discipline appropriate, there may be logical clusters of solutions for the connections between publishing and data depending on the nature of the data.

There is a role for federal agencies in coordinating across discipline boundaries (covering all funded areas) and internationally. In its October 2011 report, *Federal Engagement in Standards Activities to Address National Priorities: Background and Proposed Policy Recommendations*, the Subcommittee on Standards of the National Science and Technology Council noted that “There was agreement among respondents that the US government should continue to play the role of participant in private sector standards setting processes. There was also general agreement that the effectiveness of government participation depends on the level and consistency of involvement and commitment of resources, both staff and budgetary, to the process. Lack of coordination among agencies...was cited by many respondents as having a negative impact on government effectiveness.”

(11) What are other examples of standards development processes that were successful in producing effective standards and what characteristics of the process made these efforts successful?

The Digital Object Identifier, or DOI, is an example of a successful standard. Its development and adoption involved a multi-stakeholder, community-driven approach that solved a practical problem and provided benefit to the end-user.

(12) How could Federal agencies promote effective coordination on digital data standards with other nations and international communities?

AIP supports the recommendation of the Interagency Working Group on Digital Data (IWGDD) that an NSTC Subcommittee for digital data preservation, access, and interoperability be created. This subcommittee would then be able to provide coordination among the US funding agencies and collaborate with its international counterparts. Coordination at the national level should extend beyond

science funding agencies as relevant work is being done elsewhere within the US government (for example, the work of the Library of Congress through its National Digital Information and Infrastructure Program [NDIIP], particularly its “partnership with the National Science Foundation in 2005 to undertake a program of pioneering research to support advanced research into the long-term management of digital information”).

In addition, this subcommittee could ensure that each Federal agency is itself required to adopt and implement digital data standards developed within the global community.

Federal agencies can support conferences and other initiatives on a discipline level by funding standards and preservation work as well as pure research.

(13) What policies, practices, and standards are needed to support linking between publications and associated data?

See answer to question 9. The mechanism for linking between publications and associated data essentially exists with the digital object identifier, which is already used widely for linking between publications. The federal government could provide additional logistics and financial support for making this mechanism standard practice with respect to data and coordinating/aligning policies across federal agencies to encourage use of those standards by grantees.

Agency involvement and/or support of current initiatives such as the NISO/NFAIS Working Group on Supplementary Journal Information (www.niso.org), which is working on recommended practices for publishers who are increasingly attaching data sets as supplementary information appended to publications, would also help address some of the issues at a practical level.