

One publisher's journey through the public access debate¹

H. Frederick Dylla

Executive Director and CEO, American Institute of Physics, College Park, MD 20740, USA
Tel.: +301 209 3131; Fax: +301 209 3133; E-mail: Dylla@aip.org.

Abstract. This paper presents the content of my closing address given at the Academic Publishing in Europe 2012 Conference. I share my perspective of the public access debate, as CEO of American Institute of Physics, a medium-size scientific publisher and my observations on our industry's most important customers – the libraries. The origin of the often contentious public access debate can be traced back to a worthy goal shared by all stakeholders: the expansion of access to and broad use of scholarly publications. Starting with principles and recommendations set forth in the 2010 Scholarly Publishing Roundtable Report, I outline a productive and pragmatic path forward and identify appropriate and cost-effective options for expanding access. Furthermore, I review the major elements of public access policy development in the US since 2005, leading up to January 2012, a year after President Obama signed into public law the America COMPETES Reauthorization Act of 2010. This essay has been expanded to include additional information, covering the issue through September 2012 (submission date of this article), and addresses related government initiatives that appeared in the UK and the European Union.

Keywords: Public access, open access, scholarly publishing, publishing policy

1. Introduction

For the past three years (2010–2012), I have had the privilege of presenting the closing address for the Academic Publishing in Europe (APE) conference. My talks primarily concentrated on reviewing relevant events in the US, particularly concerning the development and implementation of policies affecting public access to scholarly publications from the perspective of publishers and funding agencies. The conference organizers felt that there were many useful connections between events in the US and in Europe – especially considering the global nature of the scholarly publishing enterprise. My 2012 APE conference address was a two-part presentation.

First, I shared my perspective on the public access debate after having served for five years as the CEO of the American Institute of Physics, a medium-size scientific publisher, and my observations on our industry's most important customers – the libraries. The origin of the often contentious public access debate can be traced back to a worthy goal shared by all stakeholders: the expansion of access to and broad use of scholarly publications. Starting with principles and recommendations set forth in the 2010 *Scholarly Publishing Roundtable Report* [5], my talk outlined a productive and pragmatic path forward and identified appropriate and cost-effective options for expanding access.

Second, I reviewed the major elements of public access policy development in the US since 2005, leading up to January 2012, a year after President Obama signed into public law, the *America COMPETES*

¹Updated to include a policy summary from January 2012–September 2012.

Reauthorization Act of 2010. This essay includes additional information, covering the issue through September 2012 (submission date of this article), and addresses related government initiatives that appeared in the UK and the European Union.

2. My observations as an STM publisher over the last five years (2007–2012)

In 2007, I had the honor of assuming the position of Executive Director of the American Institute of Physics (AIP). I came to the opportunity with significant familiarity with the submission and input side of scholarly publishing, but with very little experience with the business side of the enterprise. AIP was established in 1931 to publish journals for its founding Member Societies and soon developed an important set of its own journals. The organization also began to support a number of outreach programs with its journal income, including a sizeable library and archive for the history of science and technology.

During the three decades prior to joining AIP, I served as a practicing physicist in two US Department of Energy national laboratories, with associated academic positions. My experience as an active researcher led to numerous opportunities to author or co-author articles for a range of scientific and technical journals, with the attendant task of frequently serving as a reviewer. I had also previously served on the editorial boards of several journals and monograph series and as president of one of AIP's Member Societies – experiences which helped me to develop a better sense of the importance of journals published by a professional society. Suffice it to say, except for realizing the obvious importance of the journal income for the operations of a journal-publishing society, my exposure to the business or policy side of the scholarly publishing enterprise was limited.

My new position at AIP put me on a necessary crash course to immediately correct this limitation. In doing so, I quickly drew two conclusions about the business side of STM (scientific–technical–medical) publishing. First, the STM publishing enterprise is large and diverse, and the industry's products – including the production, dissemination and archiving of peer-reviewed scholarly articles – are absolutely essential for the advancement of scientific scholarship. In 2009, the STM publishing market published nearly 1.5 million articles from over 25,000 unique titles, from over 2,000 different publishers. Second, in the discourse over public access policy over the past decade there has been an unfortunate rift developing between scholarly publishers and their most important customers, the librarians.

Journal publishers and librarians have been linked in a collegial, mutually beneficial and necessary business relationship over most of the 350-plus year history of the journal. Through most of this period, there has been no significant evolution in the structures of the journal and the standard business model – subscriptions sales to libraries. Only since the transition to online products in the last two decades have we seen, by necessity, a rapid adaptation to new delivery methods via the online platform. Subsequently, the industry has moved to a near continuous evolution of new formats that allow a journal to be bundled with enhanced content and new services derived from technology (e.g. reference linking). A combination of factors including the constant growth in the volume of research output, journal pricing, library budgets falling behind rapidly increasing institutional research budgets, and a very genuine concern for increasing access beyond the research community for publicly funded research, have all contributed to a widening rift between these two traditionally aligned partners: publishers and librarians.

This substantial disconnect between what I saw as traditional publishing allies was troubling to me on many levels, especially against a backdrop of fond memories of working with librarians as an active researcher for so many years. I asked myself, how long can an industry stay healthy if the operable business model is a considerable and ongoing irritant to so many of its most important customers? It

is clearly in the interest of both communities to work together to repair the relationship for the sake of the authors and readers, whom we both serve, and to provide the best value to the various funders that underwrite research, scholarship and the libraries themselves.

3. The US discourse on public access

Early in my tenure at AIP, I decided that I would devote my personal energy to help find mutually satisfactory solutions that would help repair the breach. By 2009, it was clear to me that efforts to advance the debate were being seriously hampered because of polarizing comments launched by both sides. As in many debates, there was more heat than light, i.e. very little data were being applied to understanding the problem in order to catalyze constructive solutions. I do believe that all parties involved in this discussion (the funding agencies, librarians and publishers) have made some progress since then. Before describing that progress and next steps, however, I will first outline what I believe is the negative impact on the public access debate caused by very poor message crafting by publishers.

Simply put, STM publishers have shown a lack of savvy in making our case in public fora. I am struck by the irony of this particular weakness set against our general mission to distill and present clear, well-reasoned communications products. I can only surmise that this deficiency stems from some sense of entitlement acquired during the long gestation period, in the print era, when the presumption may have been that because our products sell themselves, owing to their value to the scholarly community, the industry invested little time and effort promoting the value of its journal products. The publishing landscape has clearly changed in the internet era and there are increasing demands for free information on the web. Related to this, there is a perception that since a publisher's incremental cost for providing a digital copy of a journal to the n th reader is negligible, the perceived total costs for producing online journals have become mistakenly considerably devalued.

Generally speaking, scholarly publishers have not done a good job of presenting three key issues for our customers and stakeholders concerning public access: (1) the huge increase in accessibility that online access has already provided to our readers; (2) promoting gold open access as a viable business model for some segments of our market and carefully defining these markets; (3) quantifying the "access" problem on all sides of the access debate.

We publishers have not effectively publicized the successful implementation of innovative test-piloted solutions (some in collaboration with funding agencies) to meet shared goals for access that may have helped cool the over-heated rhetoric on public access. I provide more information on some of these efforts later in this paper. We have put forward a considerable number of viable solution-based models (see Table 1) including: providing access to journal content through a low-cost article-rental service, free access to developing countries, prescribing an effective means of providing medical articles to patients and patient caregivers, and providing free access to public libraries and public high schools.

In 2010 the American Physical Society (an AIP Member Society) pioneered an effort to provide free online access to its journals to US public libraries, to better reach previously underserved groups.

Table 1

Publisher-initiated solution models for increasing access to journal content

Article access	Low-cost rental services; free access to developing countries; provision of medical articles to patients and patient caregivers
Public and school libraries	Pioneered in 2010 by the American Physical Society to provide free online access via US public libraries and later extended to public high school libraries

In 2011, the offer was extended to public high school libraries. To date, 586 public libraries and 210 high school libraries are participating. AIP and other AIP Member Societies plan to follow suit. The UK Publisher's Association successfully championed this model for inclusion in the Finch Commission recommendations.

As a consequence of our messaging failure, our pronouncements in the discourse on public access have been largely reactive rather than proactive, resulting in a very bad image problem for publishers in this debate. In the world of the 24-7 news cycle and politically driven sound bites, publishers have yet to come up with a simple response that counters the one-liner widely promoted by staunch open access advocates: "The public pays for the research; they should have access to the results". The crux of my response to this tag line is: "Of course, but freely available does not mean free of cost".

An appropriate response to this discourse is to first define in simple terms the value of scholarly publications for our diverse audiences. Moreover, we must demonstrate that the industry has put forward serious new proposals for, and have in fact deployed, a number of viable options for providing increased access. We should emphasize that it is far from a "one-size fits all" approach for either the products or access to these products.

My experience in the debate on public access, first as an observer and now as a participant, has given me a new perspective on our librarian customers – at least those who have been in the forefront of the access debate. I think it is clear that nearly every research librarian is feeling the weight of significant budget pressures and we in the publishing community need to help strengthen their collective voices as they negotiate for appropriate funding from their institutions. It is well documented that library budgets have not scaled up with the increasing amount of available journal content. Journal content continues its 350-year growth rate at three percent per annum, research budgets since World War II on average have grown even faster, but library budgets have not kept pace [9].

The question could fairly be posed: What happens when a major fraction of journals switch from a subscription-payments-to-libraries model to an author-side-payments model? I have yet to have a librarian explain to me what their role in this process would be if this transition were to happen with any significant percentage of journals or over too short a time period.

Since joining this debate, I have noticed that the generally calm and collected voices of my librarian colleagues are often overwhelmed within the debate by open access proponents in their camp who push the issue simply from an ideological perspective and consider no other factors. Yes access is good, but if quality content is compromised then our collective has taken the wrong approach.

4. Recent policy development and legislative efforts in the US

This narrative touches on some of the most impactful recent events regarding public access policy in the United States. It is by no means exhaustive. Please consult the timeline (Table 2).

4.1. NIH and PubMed Central

Negotiations between STM publishers and US National Institute of Health began in 2005 around the issue of public access to scholarly publications on an open NIH site. These negotiations were contentious and non-productive for both sides, and led to a voluntary manuscript deposit policy, where published research is posted on NIH's PubMed Central (PMC) website. In April of 2008, this voluntary policy became a mandate through legislation, requiring all scholarly articles stemming from NIH-funded research to be submitted to PMC within 12 months of publication. I believe that neither side of the negotiation

Table 2
US public access policy and legislative timeline

2005–2007	Contentious negotiations between NIH and publishers
2007	NIH Public Access Mandate passed
2008	Dueling legislation on mandates: FCRWA vs. FRPAA
2009	US House sponsored <i>Scholarly Publishing Roundtable</i>
2010–2011	America COMPETES Reauthorization Act of 2010* passed
2011	
February	OSTP forms two sub-committees (data and publications)
March	NSF's National Science Board (NSB) workshop on open data and publications
February–April	Publisher focus groups formed to work with NSF and DOE to propose and implement partnerships on access
August	OSTP pressured to show “consultation with stakeholders”
November	OSTP releases request for public comment on publications and data
November	CrossRef announces publisher–funding agency partnership for funding agency metadata (FundRef)
December	NSB releases draft report on open data
December	US House of Representatives introduces “Research Works Act” (Publishers receive negative press for opposition.)
2012	
February	House Oversight and Government Reform Committee announces “Research Works Act” will not be considered
	US House of Representatives introduces Federal Research Public Access Act of 2012 (H.R. 4004)
March	House Subcommittee on Investigations and Oversight holds hearing on “Federally Funded Research: Examining Public Access and Scholarly Publication Interests”
March	OSTP releases report, “Interagency Public Access Coordination”, includes information gleaned from the public “Request for Information” (RFI)
May	FundRef pilot program announced
June	Finch report published in the UK

*Includes many Roundtable recommendations and directive for an inclusive process but no language on mandates.

should be proud of this episode in agency-publishers relations, nor should it be a model for moving forward for other fields of research supported by other funding agencies.

Some within the research community find the convenience and breadth of PMC to be a valuable tool, but others prefer the curation of publisher websites and third-party databases more useful. According to NIH, about 2.4 million articles are currently represented on that site. However, some observers, both within the publishing industry and in other federal agencies, are concerned about both the costs of PMC and the duplicative nature of its offerings.

Many publishers are apprehensive about the prospect of harm from the NIH mandate. However, the complexities of publishing economics make it difficult to demonstrate direct evidence of harm, and indirect measures have been dismissed by NIH. It is the opinion of many supporters of the NIH mandate that its imposition has not shown definitive harm to medical publishers. This lack of evidence has strengthened support for an extension of this type of legislative solution for improving public access to other fields of research supported by the US government.

Publishers have expressed several concerns about PMC its implementation. These include:

- (1) Diversion of Traffic – There is growing evidence that readers are downloading journal content directly from the PMC site rather than linking to the publisher's version of record on their sites

[1–3]. Decreases in downloads from publisher's sites is a key factor cited by librarians as influencing their renewal decisions [4];

- (2) Piracy – There is some evidence that PMC content is being skimmed and re-packaged by unauthorized third parties for competition on the commercial market with the original content providers; and
- (3) Unnecessary Government Influence – There is concern that NIH, because of its size, often wields unnecessary influence on the development and wide acceptance of standards for publication platforms because of its independent promotion of article and author identifiers. The NIH did offer the wider community a service by promoting the adoption of NLM-DTD as a standard XML markup language, but it has resisted similar efforts to promote the widely adopted article identifier (the CrossRef DOI²), and so far remains committed to its independent author identifier (PMCID), rather than joining the new ORCID consortium.³

4.2. *FRPAA and RWA*

The proposed Federal Research Public Access Act (FRPAA), introduced in the US Congress almost every year since 2008, would extend the NIH's public release mandate to all disciplines represented by the eleven US federal agencies that fund research above \$100M per annum. In February of 2012, the FRPAA bill was reintroduced in both Houses of the US Congress with similar statutes as previous versions of the bill except that it would decrease the embargo time frame to 6 months. Also within the last year, the Research Works Act (RWA) was introduced. RWA would have prohibited the free dissemination of private research works funded by federal agencies, thus nullifying the current NIH mandate and any such future mandate. The effort failed to gain traction – ultimately congressional cosponsors agreed not to take legislative action – but it nevertheless served to intensify the debate. Although many publishers, including AIP, did not support the legislation, several important scholarly publishers did support it, and the industry as a whole received very poor publicity for the introduction of this ill-fated and poorly timed legislation.

4.3. *The Scholarly Publishing Roundtable and COMPETES*

In June 2009, the US House Science and Technology Committee convened the Scholarly Publishing Roundtable to provide recommendations to the federal government on how to accomplish public access to federally funded research. The roundtable members were drawn from the library, university administration, research and publishing communities. The core recommendation of the Roundtable report (see Ref. [5]) recognizes the responsibility of the federal government to develop policies for public access to the research results of publicly funded research. Many of the Roundtable's recommendations were incorporated in Section 103 of the America COMPETES Reauthorization Act of 2010, calling for federal funding agencies to engage key stakeholders in the development of pragmatic public access policies.

²CrossRef is a collaborative reference linking service that affects linkages through CrossRef Digital Object Identifiers (CrossRef DOI), which are tagged to article metadata supplied by the participating publishers. Available at: www.crossref.org/01company/02history.html.

³ORCID is an international not-for-profit organization to enhance the scientific discovery process and improve collaboration and the efficiency of research funding. ORCID aims to solve the name ambiguity problem by creating a registry of persistent unique identifiers for individual researchers and an open and transparent linking mechanism between ORCID, other ID schemes, and research objects such as publications, grants and patents. Available at: <http://about.orcid.org/>.

In response to COMPETES, federal agencies have engaged commercial and nonprofit publishers, universities and others, to develop sustainable approaches toward increasing public access while preserving incentives for investments in private-sector research works such as scholarly publications. In early 2011, AIP and several other scientific publishers began working directly with US Department of Energy (DOE) and the National Science Foundation (NSF) to develop agency–publisher partnerships and pilot projects for increasing access to and interoperability among agency and publisher databases. These efforts are detailed in Section 6.

4.4. OSTP report

The US Office of Science and Technology Policy (OSTP) was given responsibility for overseeing the interagency planning called for by the COMPETES Act. OSTP released a report [8] at the end of March 2012 on the progress one year after the bill's passage. OSTP did a significant amount of work on this complex topic by commissioning and managing two interagency panels as described in the report (one on access to data and one on scholarly publications). The primary content in the report refers to developments on public access from three agencies: NIH, which already implemented a public access policy for publications in 2008 by legislative mandate, and from DOE and NSF, which began development of public access policy as a result of the COMPETES statutes.

The report notes that OSTP will continue to monitor progress at the agency level and leaves open the possibility for additional reports in the future as agencies make further progress on policy development and implementation. From my standpoint, I am pleased that the focus has shifted from the NIH to DOE and NSF. As indicated in Section 6 on agency–publisher partnerships, DOE and NSF serve as models for how the other agencies can respond to the related issues of access, interoperability standards and archiving.

As a scientific publisher, I am pleased to note the following two quotations from the OSTP report:

...agencies and public commenters are cognizant of the essential role that publishers and the peer review system play in advancing the scientific enterprise. The PASP therefore set out to explore what steps could be taken to expand public access while preserving the value that publishers provide to the scientific enterprise, creating new business opportunities, and maximizing the economic and societal benefits of the Federal investment in research and the resulting publications.

It should also be stressed that it is the intention of the Administration to continue a robust dialog with the private sector and the public to ensure that policies developed will benefit the public interest and to maintain a level playing field for all interested parties.

5. Assessment of public access policy

Government initiatives designed to increase public access to scholarly publications in the US and in the EU 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.

It is the position of AIP and many other STM publishers that forcing all STM journals to adopt a pre-paid open access model or contend with mandated embargo release periods are not universal solutions for the diverse array of journals published across all the disciplines represented by government 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 or decades, such as mathematics, theoretical physics and the social sciences. Most publishers feel that 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 innovate a wide range of products and dissemination methods.

6. Developing agency–publisher partnerships

The America COMPETES Act directs US federal agencies 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 scholarly publishing continues to be a self-sustaining enterprise. 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. Moreover, they can generate new audiences for their articles, and new approaches toward linking information in ways that improve scholarship and discovery.

A most appropriate role for the federal government is to encourage federal agency–publisher partnerships, examples of which have arisen as a direct result of the COMPETES legislation. Shortly after the passage of the COMPETES legislation, a small group of STM publishers formed a working group to interact with two leading US science agencies, DOE and NSF, to propose and plan partnerships and pilot projects on the topics of concern spelled out in Section 103 of the COMPETES law. Topics included increasing access to all the products derived from publicly funded research, grantee reports, data and peer-reviewed publications, 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.

6.1. Standards and identifiers: Agency funding information

Most funding agencies currently require researchers to acknowledge in publications the support that they have received. However, there is no standard methodology on how this should be done. Consequently, agency funders find it difficult to know what publications have arisen from the research they have funded. A group of six STM publishers (including AIP) have recommended that the publishing community develop, in collaboration with funding agencies and CrossRef, a means of standardizing funder-identification information and make it available to funding agencies and the public. This initiative is based on the belief that a community-wide solution of this type will be easier and far less expensive to implement than for each agency to develop its own methodology of data mining in response to the problem. The prevailing conventional wisdom is that publishers are best positioned to provide a simple way of ensuring that journal articles are accompanied by standardized, high-quality metadata providing information about the agency, program, and even the specific grant that funded the research.

CrossRef and several major STM publishing trade associations have endorsed this proposal including the Professional and Scholarly Publishing 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. In addition, one of the participating publishers, Elsevier, has agreed to donate to the collaboration a dataset from Elsevier's Scopus database that includes funding information for millions of articles available through their ScienceDirect publication platform.

CrossRef has agreed to manage this collaboration between the publishers and, so far, four funding agencies under the name of the "FundRef" initiative.⁴ Pilot demonstration projects are expected to be launched in 2013. With the successful implementation of the FundRef project, funding agencies would have access to standard metadata from published articles through CrossRef licenses that most agencies already maintain. By displaying this information on agency websites, visitors – from the research community to the general public – could follow the link, enabled through the publication's Digital Object Identifier (DOI) to the publisher's platform where article abstracts are freely available. The Version of Record (VoR) of the full article maintained by the publishers would be made available through a variety of publisher-determined access mechanisms, including, for example, an innovative rental access model that gives the public instant access for a modest fee. More than 40 scholarly publishers are currently testing this particular access mechanism.

6.2. Standards and identifiers: DOIs for data sets and supplementary material

Globally, funded investigators are more frequently being asked to share or provide plans on how they will share with other researchers the primary data, samples, physical collections, and other supporting materials created or generated 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.

To facilitate the establishment of data policies, publishers would draw on their experience with initiatives such as Opportunities for Data Exchange (see www.alliancepermanentaccess.org/current-projects/ode), designed to gather and promote best practices on the way scientific data are treated. Similarly tasked is CODATA, an interdisciplinary scientific committee of the International Council for Science (ICSU). The goal of the relatively new ICSU World Data System (WDS, www.icsu-wds.org) is 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. (An interim report is available: www.niso.org.)

The NSF Directorate for Math and Physical Sciences has recently funded a proposal submitted by AIP and the American Astronomical Society to establish a pilot project to link the data behind figures and tables with publications, including developing author protocols for identifying, and archiving the high value data that is associated with peer-reviewed publications.

⁴FundRef is a collaborative pilot project of scholarly publishers and funding agencies, facilitated by CrossRef, to provide a standard way of reporting funding sources for published scholarly research. See www.crossref.org/fundref.

6.3. *Standards and identifiers: Author disambiguation*

Author-name ambiguity and funding-attribution vagueness are persistent, critical problems endemic in the scholarly research ecosystem. STM publishers are encouraging funding agencies to work in collaboration with publishers as well as universities and corporations from around the world to eliminate this problem through the ORCID partnership. 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. (See footnote 2.) 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. ORCID is governed by representatives from a broad cross section of stakeholders, including publishers, library organizations, research institutions and funding agencies.

6.4. *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. Publishers are committed to managing content in modern digital formats to ensure that users gain maximum benefits. Accordingly, we are in discussions 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 to identify relevant information and facilitate its extraction and analysis. The Publishing Research Consortium recently completed a study [6] on article-level content mining based on a broad survey of ongoing or planned activities among nearly 30 STM publishers or associations.

Data-mining pilot projects would 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.

6.5. *Pilot projects: Sponsored access to published research*

The 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 VoR.

NSF and the Max Planck Society [7] have been leading a discussion with a group of physics publishers to set up a pilot to answer the following questions dealing with the cost, benefits, range of applicability and sustainability of the 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 open access in the aggregate and on a per-article basis?
- What is the most effective method to provide open access funding for authors?
- 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?
- How can agencies best administer a gold open access program?
- Does open access offer agencies new opportunities to showcase the productivity of their funding activities to the public that funded the relevant research?

6.6. *Pilot projects: Linking to/from research reports*

Pilot projects on linking to/from research reports would try to determine whether and/or how publisher content derived from agency funded research could be mapped against agency research reports and other related content. Specifically, such pilots 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 possibly full-text versions linked to the Version of Record of articles connected to a particular research report or funded project.

If successful, this would result in interoperability between online agency content and publisher platforms. This is of interest to 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 (if full text versions were not available by other means),
- 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 (<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.

In mid-2011, a pilot project of this type was initiated by Wiley working with DOE-OSTI. 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 their existing search protocols, OSTI would make the journal publisher's

full text searchable, with the intent to make citations of DOE-funded journal articles available in the search and retrieval applications operated by OSTI. By mid-2012, this pilot project with DOE-OSTI was being extended to other STM publishers.

7. Actions in UK and EU

In the summer of 2012, there were important announcements out of Europe on the issue of public access that caught the attention of observers in the US and worldwide. The UK government released its official comments [10] on the Finch report [11], a yearlong study of how to best provide wide-ranging access to publications that are underwritten by UK funding agencies. The report panel was chaired by Dame Janet Finch, a professor of sociology at the University of Manchester. The UK government largely accepted the recommendations contained in the Finch report.

The Finch Group, commissioned by David Willetts, the UK minister of state for universities and science, is similar to the effort in the United States with the Scholarly Publishing Roundtable, on which I had the privilege of serving in 2009.

Among the Finch recommendations accepted by the UK Government:

- Moving to deliver open access through a “gold” model, where article processing charges are paid upfront to cover the cost of publication;
- Implementing walk-in rights for the general public, allowing free access to global research publications owned by members of the UK Publishers Association via public libraries;
- Extending the licensing of access by universities to high technology businesses for a nominal charge; and,
- Beginning April 2013, science articles must be made freely available by one of two routes: preferably upon publication via the “gold” model or within six months of publication on a publically accessible repository (although Finch recommended more flexibility on embargo lengths) – if the articles come from research paid for by one of the UK’s seven government-funded grant agencies, the Research Councils UK (RCUK), who annually invest £2.8 billion (\$4.4 billion US).

RCUK and the Higher Education Funding Council for England (HEFCE) also both signed off on the major points contained in the Finch report. It is important to note the UK Research Councils are planning to provide block grants to research institutions to support author payments. Also in July 2012, the European Commission (EC) adopted two communications and one recommendation relating to public access, proposed for inclusion under an overarching, 7-year plan for EC research funding called Horizon 2020. The EC will establish public access to scientific publications as a tenet of Horizon 2020, the research and innovation funding program for 2014–2020, but there are several significant steps that have to be taken before this plan is actually codified. At this stage of the process, there are no EC provisions for funding open access through a prescribed “gold” model, and it should be noted the EC process is to date a far less collaborative effort than was the Finch group. Of course, “single payer” solutions such as proposed by the Finch group and the EC recommendations are not as adaptable to the more complicated research infrastructure that exists in the US, which is why AIP has and will continue to advocate a mixture of solutions for public access in the United States (e.g. rental models, linking between agency and publisher sites, public library access, and others) as our preferred approach for the evolution of public access.

Especially given the highly polarized political climate in the US surrounding the public access debate, it was particularly gratifying for me to see my European colleagues come together with their respective

divergent viewpoints on this issue and present unified recommendations for UK government adoption, which they believe will work for them. The Finch conferees showed deftness and resolve in their delicate negotiations among funding agencies, publishers, universities, librarians, and researchers. The timing of this report will influence the ongoing deliberations in the US on how to expand access to scholarly publications and other research results from public funding to the widest possible audience.

8. Conclusion

Fundamentally, I believe that a uniform public access policy or a mandate covering all 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.

AIP believes that governments should support and encourage access to scholarly publications via mutually beneficial partnerships with publishers that would foster economic development and maximize the productivity of the scientific enterprise. The best examples of this are 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, that offer a pragmatic and productive route to success in broadening public access to 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 US scholarly publishing community that new legislation is not needed at this time. Post-COMPETES evidence shows that there has been real progress on the interrelated issues of access and interoperability among public and private information platforms and databases.

While AIP strongly supports the posting of grantee reports on agencies' public websites – and I believe that the public is justified in their expectation to see publications and associated data derived from the research – I am compelled to consider the costs of publishing a typical article in an AIP journal, \$2500 on average. Other journals may have business models that make this cost more or less expensive. The point is this expense must be recovered through subscriptions, pay-per-view fees, or in the case of some journals, page charges, other author fees or some other modality. The present mixed array of business models would need to be maintained throughout the transition to wider public access to ensure the sustainability of the scholarly publishing enterprise.

In my opinion, the public access debate has been a huge distraction for all sides. It has focused on access to scholarly publications for the obvious reason that a journal article is a tangible and highly valuable research-based product. I believe, in the face of increasing budget pressures faced by our librarian colleagues, we in the publishing community could be more vocally supportive of their negotiations with their institutions for increased funding for library programs.

Finally, although many of the recommendations in the UK's Finch report differ somewhat from AIP's views, the effort should be applauded for its inclusivity, which led to the commission's united recommendations for a path forward. Significantly, Finch's recommendations include models similar to US funding agency–publisher partnership pilots underway, acknowledge the value STM publishers provide to the research community, and offer mechanisms to cover basic costs for journal publishing. I believe the entire community of stakeholders in this debate will continue to identify and ultimately find mutually satisfactory solutions to provide for the public's right to access the results of publicly-funded research.

And then, we can move on to the important tasks which have always aligned the publishing and library communities – including how we collectively disseminate, catalog and archive scholarly works.

References

- [1] R. Apodaca, Depth-First blog post: ACS responds to request for information on the new NIH public access policy, 2012, available at: <http://depth-first.com/articles/2008/06/03/acs-responds-to-request-for-information-on-the-new-nih-public-access-policy/>.
- [2] P. Davis, The effect of public deposit of scientific articles on readership, *The Physiologist* (in production for October 2012), available at: www.the-aps.org/mm/Publications/Journals/Physiologist.
- [3] M. Frank, “Public access” to scientific literature?, *The FASEB Journal* **26** (2012), 1767–1770.
- [4] Publishers Communications Group, A view of aggregators from the librarian's perspective (2008), available at: www.pcgplus.com/pdfs/aggregators_librarians_perspective.pdf.
- [5] Report and Recommendations of the Scholarly Publishing Roundtable, January 2010, available at: www.aau.edu/WorkArea/showcontent.aspx?id=10044.
- [6] E. Smit et al., Journal article mining: A research study into practices, policies, plans ... and promises, Commissioned by Publishing Research Consortium, 2011, available at: publishingresearch.net/documents/PRCSmitJAMreport20June2011VersionofRecord.pdf
- [7] T.S. Statler, NSF's initiatives in expanding public access to research results, presentation at the 2012 AIP Assembly of Society Officers, 2012, available at: <http://www.aip.org/aip/assembly/march12/index.html>.
- [8] US Office of Science and Technology Policy, Interagency public access coordination: A report to congress on the coordination of policies related to the dissemination and long-term stewardship of the results of federally funded scientific research, 2012, available at: www.whitehouse.gov/sites/default/files/microsites/ostp/public_access-final.pdf.
- [9] M. Ware and M. Mabe, STM report: An overview of scientific and scholarly journals, September 2009, available at: www.stm-assoc.org/industry-statistics/.
- [10] D. Willetts, Government Response to the Finch Group Report: “Accessibility, sustainability, excellence: How to expand access to research publications”, 2012, available at: www.bis.gov.uk/assets/biscore/science/docs/l/12-975-letter-government-response-to-finch-report-research-publications.pdf.
- [11] Working Group on Expanding Access to Published Research Findings, Accessibility, sustainability, excellence: How to expand access to research publications, 2012, available at: www.researchinfonet.org/publish/finch/.