



Allan R. Adler

Vice President, Legal & Governmental Affairs

January 12, 2012

Re: FR Doc. 2011–28621

On behalf of the Professional and Scholarly Publishing Division of the Association of American Publishers (“AAP/PSP”), I am pleased to respond to the Office of Science and Technology Policy’s (“OSTP”) November 3, 2011 Request for Information (“RFI”) regarding “Public Access to Digital Data Resulting from Federally Funded Research.”

AAP/PSP members publish the vast majority of materials used in the U.S. by scholars and professionals in science, medicine, technology, business, law, reference, social science and the humanities, and they include the worldwide disseminators, archivists and shapers of the public record on scientific research via print and electronic means. They include non-profit professional societies, commercial publishers and university presses that create books, journals, computer software, databases and electronic products in virtually all areas of human inquiry and activity.

Collectively, AAP/PSP members represent tens of thousands of publishing employees, professional individuals, editors and authors throughout the country who regularly contribute to the advancement of American science, learning, culture and innovation. They comprise the bulk of an \$8 billion commercial and non-profit publishing industry that contributes significantly to the U.S. economy and enhances the U.S. balance of trade by at least \$3.5 billion annually.

For AAP/PSP members that publish scientific journals and other peer-reviewed scholarly publications, the primary goal of their publishing activity is to disseminate information and provide access by providing a high quality and user-friendly digital environment in which to discover, analyze and link to the latest breakthroughs and developments in scientific and other scholarly research. In particular, publishers of scientific journals have, for more than 100 years, played an integral role in building and documenting the unrivalled U.S. scientific research enterprise. In addition to their efforts to disseminate publications that report and analyze the

latest research, they also have considerable experience and investment in digital technology, metadata standards and tools to help users understand and manipulate data. This makes publishers uniquely positioned to help the Federal Government in expand public access to digital data, ensure the long-term stewardship and discoverability of data and support the innovation and economic development that is derived from scholarly advancements.

It is worth mentioning at the outset that, in contrast to peer-reviewed publications, which are not the “result” of federally-funded research, digital data does often directly result from research funded by the government. Research and publication are unique creative acts. Publishers support better discoverability and reuse of scholarly data and are pleased that OSTP has recognized the distinction between data and peer-reviewed publications in this RFI. The dissemination of information is an area of publishers’ professional expertise, and data access policies potentially impact AAP/PSP members not only as publishers of peer-reviewed scientific journals, but also as disseminators of information whose innovative products and services enhance and add value to taxpayer-funded research activities and are expected to do so in the future.

It is with this view that the attached comments and recommendations have been submitted on behalf of AAP/PSP in the hope that they will help to facilitate the successful development of sustainable and effective policies on public access to data that are consistent with the Administration’s “Open Government” framework¹ and embodies a spirit of collaboration in recognition of the intellectual property rights and private investments of publishers as key stakeholders in these matters.

General Recommendations

Scholarly publishers have long served as integral hubs of the America’s research enterprise, validating research through the peer review process, producing a scientific record and facilitating scholarly communication through dissemination and preservation of scientific literature, including the presentation and long-term stewardship of digital data. The primary goal of publishing is to facilitate the widest possible dissemination of the information that publishers create. In the digital age, publishers have invested significantly to enhance the discoverability, public access to and the utility of research data, particularly for the scientific, technological, engineering, social science and medical communities: expanding accessibility,

¹ As articulated in Memorandum for the Heads of Executive Departments and Agencies on Transparency and Open Government (January 21, 2009), available at http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment and Memorandum for the Heads of Executive Departments and Agencies on Open Government Directive available at <http://www.whitehouse.gov/open/documents/open-government-directive>

improving interoperability and fuelling innovation. Publisher investments have created digital platforms with the latest and continually evolving Web capabilities, providing researchers with faster and more robust delivery of scholarly information, new ways to present data and research findings and links that enable information to be found and navigated with ease. Publishers have improved interoperability through new metadata standards and pilot projects, which are driving innovation and providing for better information discovery and expanded use of research results. As long as the government does not diminish incentives for creative publication, publishers will continue to provide tools that enhance innovative reuse and discovery of research information. Publishers look forward to continuing a positive collaboration to enhance science and innovation in the United States, and welcome any partnership with the Administration to harness the power and potential of technology and innovation to spur long-term economic growth and provide cutting-edge solutions to support domestic priorities.

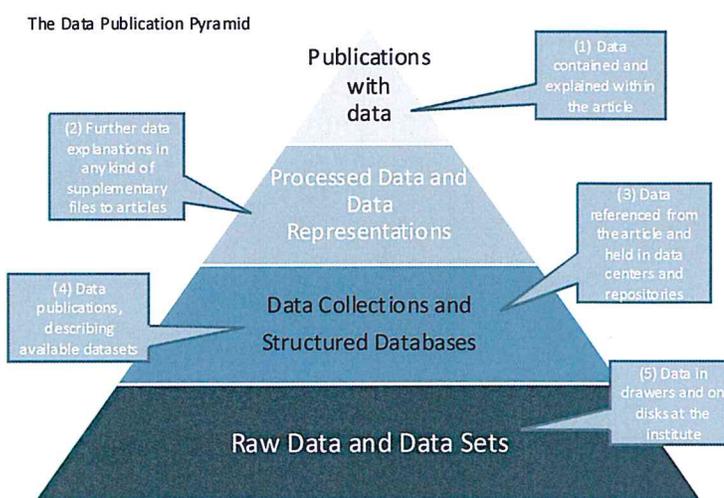
At the same time, the experience of publishers informs an understanding that significant challenges remain to wider public access to and long-term stewardship of digital data. Unlike publications which can be counted by the number of pages, datasets can be measured in mega-, giga-, or even terabytes of data. Unlike publications, which undergo peer-review and editing, extensive formatting and tagging, data files come in a variety of forms, formats and levels of quality and validation. It is difficult to control for quality and consistency. While publication incentives have been embedded in the academic and research process, incentives for complete tagging of individual datasets are limited. There is no consistent approach to presentation, standards for which data should be preserved or overall management and no consistent responsibility for data storage, tagging and dissemination. In addition, the significant costs of storage, distribution bandwidth and overall management and curation must be addressed. These inconsistencies and unknowns must be addressed in a collaborative manner among all stakeholders to promote the development of robust, sustainable and flexible standards that meet the needs of users at all levels. Publishers stand ready to lend their expertise to such a collaborative process to provide value to the research community and to the taxpayer.

The government has a responsibility to provide broad access to the digital data that results from federally-funded research, in contrast to the peer-reviewed publications that contain significant value-added beyond the federal investment. At the same time, the government should not invest funding or energy to recreate what is already being achieved by the private sector. The government's best approach is to leverage public-private collaborations, ensuring the continued innovations in publishing that contribute to the progress of science, allow innovation to flourish and help grow the American economy. A federal role in expanding access to and the preservation of digital data could include partnering with the scholarly community for the identification of standards and best practices for the interoperability of data repositories; creating clear rules for citation, modification and privacy; improving links between data, research grant reports and peer-reviewed publications; facilitating cyber infrastructure

and collaboration within and between federal agencies; and advancing policies and funding to ensure the long-term sustainability of data archives. Public access policies should be developed through voluntary collaborations with nongovernmental stakeholders, including researchers and publishers, university administrators, librarians and the public.

OSTP could learn from initiatives already underway to standardize metadata and provide links between sources of research information. In response to the several questions posed in the RFI asking for best practices or suggested approaches to expanding access, managing data, minimizing compliance costs and other policy questions, AAP/PSP encourages OSTP to review these voluntary projects, develop relationships with groups engaged with the issue and encourage the continued evolution of programs that are working to improve data stewardship and public access to data. These include CrossRef (www.crossref.org), DataCite (www.datacite.org), Opportunities for Data Exchange (www.ode-project.edu), APARSEN (www.alliancepermanentaccess.org/index.php/current-projects/aparsen/) and the NISO/NFAIS Working Group on Supplementary Journal Information (www.niso.org), among others. Such collaborative approaches provide the best way forward towards broad access to and preservation of digital data.

It is critical that the federal government continue to distinguish between data and various types of presentation of data and preserve and respect intellectual property protection and copyright ownership as appropriate. The Data Publications Pyramid displayed here,² derived from open science pioneer Jim Gray's e-science pyramid, provides a model for understanding how research data can be presented in a variety of ways with increasing levels of curation and analysis. Federal policies should take into account the differences between information products at different levels of the pyramid and work with all stakeholders, including primary researchers, secondary researchers, publishers, libraries and data centers, to create clear rules and protocols for the



² As appearing in the October 17, 2011 *Report on Integration of Data and Publications*, a report of Opportunities for Data Exchange which brings together stakeholders including researchers, publishers, libraries and data centers to support a more connected and integrated scholarly record. Full report available at http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf

sharing of data. A collaborative approach will ensure that the needs of each stakeholder group are addressed and that the progress of science is not impeded. In particular, the need to expand incentives for providing broad and timely access to new data must be balanced with the need to preserve incentives for researchers to interpret and analyze their results through curation and peer-reviewed publication.

Rather than imposing an inflexible mandate, federal policies should focus on supporting and encouraging the development of cyber infrastructure, standards for the structure of data and metadata, navigation tools and applications to achieve discoverability and interoperability and ensuring appropriate and sustainable funding for innovation and long-term stewardship. These policies should be developed in collaboration with all key stakeholders involved in the presentation, analysis, deposit, storage and preservation of data.

Responses to specific questions:

Many of the questions posed in the RFI ask for the best approach to specific issues involved in access, interoperability and preservation of digital data. The general response above offers AAP/PSP's recommendation for the government's approach: encouraging OSTP to proceed with a sensible, flexible and careful approach, and to learn from and leverage the experience of successful pilot projects and public-private collaborations. Below I add additional comments in response to questions 2, 5, 8, 9, 11 and 13, where publishers, with their long experience in presenting, showcasing and archiving data, have additional information that may be of use to OSTP.

(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?

Intellectual property, copyright and patents are critical features of the scientific and scholarly research world that incentivize innovation and the development of value-added products related to scholarly discovery. Any access policy must comply with current law and provide clear rules for sharing, citation and acknowledgement of any modifications. Federal agencies should recognize that there are different levels of data publication, as described in the Data Publications Pyramid above and craft policies that appropriately protect the value-added in curated collections, processed data and publications. Additionally, appropriate policies, developed in consultation with all stakeholders, should be developed to protect the rights of the creator of data to publish analysis and interpretation of the data in peer-reviewed publishing.

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

AAP/PSP 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.” A critical component of any policy needs to be collaboration with researchers, publishers, librarians, universities and research institutions in an interconnected system based on community needs, standards and best practices. Each stakeholder community can contribute their expertise and ensure the creation of data management policies that reflect the different practices of individual research communities.

The involvement of each stakeholder will ensure the preservation of incentives for innovation and help improve information sharing and training within each stakeholder community. Stakeholders can help develop clear standards and guidelines for the availability of research data, certification and auditing of data repositories and metadata standards, which respect each community’s standards and practices, working together to create universal policies that work for all communities. Stakeholder input is also important for the integrity of the scholarly record, including the creation of links between datasets and the scholarly publications that analyze and interpret the data. Finally, stakeholder input is necessary to incentivize the deposit of datasets and minimize the administrative burden on researchers, libraries and publishers.

(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?

Over the past decade, publishers have invested significantly in providing new services and interactive tools to enable innovation and make information more useful and discoverable. Publishers will continue to innovate in ways to present data to advance science and grow the economy in partnership with researchers and industrial users of data.

The most important step that agencies could take to promote broader use of data is to promote a comprehensive framework for reliable digital data preservation, access and interoperability through the promotion of standards and clear rules developed by the scholarly community. Agencies could also support pilot projects, data curation programs and interpretation initiatives for the relevant scholarly disciplines. Finally, agencies could use their web presence to provide a clearinghouse to the data they hold or which is funded by their grants.

(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?

The scholarly community already has a robust attribution and credit system with respect to peer-reviewed publication. This could be leveraged in a bi-directional manner by linking between datasets and publications on the one hand, and a publishing requirement that all data which informs the analysis and conclusions of a peer-reviewed publication be cited according to community standards on the other.

The federal government could help by promoting those standards and provide clear rules for the citation of datasets and acknowledgement of modifications to source data. They should also promote unique and persistent identifiers for data and disambiguate researcher, institution and funder information in metadata. Over the past decade, publishers developed the Digital Object Identifier (DOI), a unique code for each piece of content in a scholarly publication, and similar identifiers are being developed by DataCite³ for data (www.datacite.org). The work of DataCite, CrossRef and DOE's Data ID Service should be leveraged to ensure data is appropriately archived and recognized as primary research output.

(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 (DOI), a unique code for each piece of content in a scholarly publication, was developed and adopted through a multi-stakeholder, community-driven approach. It is successful because the standard evolved in response to a real problem in scholarly communication and is providing practical benefits to users of published research.

Digital data standards are newer and still evolving. OSTP should learn from ongoing initiatives that are working to address real problems in collaborative public-private partnerships with stakeholders, such as:

- NISO/NFAIS Supplemental Journal Materials Working Group (<http://www.niso.org/workrooms/supplemental>), which is working to address technical issues surrounding the definition, publication and linking of journal articles and

³ DataCite is a non-profit organization whose aims are to establish 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. DataCite is currently engaged in the process of helping researchers find, identify, and cite research datasets; providing persistent identifiers for datasets, workflows and standards for data publication; and enabling research articles to be linked to the underlying data. To achieve these goals, they are currently working primarily with organizations that host data, such as data centers and libraries.

supplemental materials, including data, as well as archiving, preservation and migration of different file formats;

- [DataCite](http://datacite.org/) (<http://datacite.org/>), which is working collaboratively address the challenges of making research data visible and accessible;
- [APARSEN](http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen/) (<http://www.alliancepermanentaccess.org/index.php/current-projects/aparsen/>), which is working through a collaborative group of more thirty research institutes, national libraries, IT providers and research funders to create a Network-of-Excellence on digital preservation;
- [Opportunities for Data Exchange \(ODE, \[www.ode-project.eu\]\(http://www.ode-project.eu\)\)](http://www.ode-project.eu), which is working to promote best practices around the way scientific data are treated;⁴
- [PARSE.insight](http://www.parse-insight.eu/) (<http://www.parse-insight.eu/>), which published a roadmap and recommendations⁵ for long-term accessibility and usability of scientific digital information in Europe; and
- [CoData](http://www.codata.org/) (<http://www.codata.org/>), an interdisciplinary scientific committee of the International Council for Science Unions (ICSU) currently working on an initiative for a World Data System.

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

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. “ AAP/PSP fully endorses this role for OSTP and the federal agencies.

Strong incentives and guidance for coordinating and aligning policies among federal agencies are critical, but much of the infrastructure is already available to create these linkages. The Digital Object Identifier (DOI) is already used to provide links between publications. With appropriate policies (see question 9) and metadata standardization (see, for example, the CrossRef initiative), the DOI could enable links between publications and associated data.

⁴ ODE’s *Report on Integration of Data and Publications* is available at <http://www.alliancepermanentaccess.org/index.php/current-projects/ode/outputs/>

⁵ The *Insight into Digital Preservation of Research Output* report is available at http://www.parse-insight.eu/downloads/PARSE-Insight_D3-6_InsightReport.pdf and the *Science Data Infrastructure Roadmap* is available at http://www.parse-insight.eu/downloads/PARSE-Insight_D2-2_Roadmap.pdf

As mentioned above, AAP/PSP recommends that agencies coordinate with and support ongoing initiatives to address the technical and practical issues involved in linking data and publications. For example, the NISO/NFAIS Supplemental Journal Materials Working Group (<http://www.niso.org/workrooms/supplemental>) is preparing draft recommendations on how publishers can best attach data as supplementary information to a peer-reviewed publication. As mentioned earlier, DataCite is similarly working to determine best practices for data archiving and metadata standards which would be critical to providing the links needed.

Sincerely,

A handwritten signature in blue ink that reads "Allan R. Adler". The signature is written in a cursive style.

Allan R. Adler
VP, Legal & Governmental Affairs