



Interagency Working Group on Digital Data
National Science and Technology Council
Executive Office of the President
Washington, DC 20502

12 January 2012

Response to Request for Information: Public Access to Digital Data Resulting From
Federally Funded Scientific Research (76 Fed. Reg., 68517, 4 Nov. 2011)

To Whom It May Concern:

With over 400,000 members in over 160 countries world-wide, IEEE is the world's largest professional association dedicated to advancing technological innovation and excellence for the benefit of humanity. In addition to our conferences, standards and other activities, IEEE publishes more than 150 transactions, journals and magazines, which represent more than 30% of the world's annually published literature in electrotechnology, computing and related fields.

IEEE is a strong supporter of public access to government-funded research results within a context that protects and advances other important societal interests inherent in scholarly publishing, including peer review and reserving the intellectual property rights of authors and publishers. IEEE's long-standing policies permit authors to "self-archive," i.e., to post their articles either on their personal web sites or their employers' web sites, consistent with open access practices. IEEE is also experimenting with several open access business models with the goal of supporting sustainable open-access publishing. Our views were outlined in greater detail in a 2007 IEEE Position Statement on "Scholarly Publishing", which is available on-line at:

http://www.ieee.org/documents/IEEE_Publishing_Principles.pdf.

IEEE has responded to the Office of Science and Technology Policy RFI on Public Access to Scholarly Publications, issued in parallel with this RFI on Public Access to Digital Data. Of course, there are qualitative differences between data sets and research articles that would lead IEEE to separate its positions in these two responses:

- The scholarly article is an author's expression of ideas with significant value added by the publisher, primarily in the form of peer review, but also in copy editing and formatting, and in other ways as well. Data sets, in and of themselves, are not normally considered part of a research article, but rather supplemental material sometimes made available with an article. However, advances in science and

technology in recent years show the immense potential of well-curated data repositories, and it is appropriate to understand necessary distinctions between published research articles and the data that underlie them.

- Science and engineering benefit from a free exchange of data that allows researchers to use data in the replication of primary research or to build upon prior research. Curation of data sets on behalf of scientific and engineering communities is a valuable service. IEEE supports public initiatives that enable researchers to make their data available to such repositories. To date, scholarly publishers have not made an investment in data archives that would be comparable to their collected volumes of published research articles.
- Although case law is developing for copyright in compiled databases, data are generally accepted as non-copyrightable facts, and therefore more amenable to free exchange without encumbrance of copyright. This is not to assume copyright does not or will not play a role in data management, only that a tradition does not exist in the same way that copyright is clearly established for collected research articles.

Ongoing maintenance of a growing number of repositories of data sets will eventually require a financial model to assure sustained support. Such models exist for research publications, including those made universally available via open access. Private-sector companies have readily exploited databases on a sustained and profit-making basis. Public databases supported by government funding are familiar, and institutional repositories are gaining experience in archiving data sets. In some cases, not-for-profit organizations have made data sets publicly available for free, while charging fees for value-added services to defray costs of archive maintenance.

IEEE strongly supports formal collaboration among stakeholders in public-private working groups. As with the policy-making process for public access to scholarly publications, we would be pleased to provide representatives to serve on collaborative panels convened to deal with these issues.

Please see detailed responses to OSTP questions following this letter.

Sincerely,

A handwritten signature in blue ink that reads "Gordon Day". The signature is written in a cursive, flowing style.

Gordon Day
IEEE President and CEO

With respect to the specific questions posed in the Request for Information, IEEE is pleased to provide the following input:

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?

The US Federal government may play a most beneficial role in partnership with standards-setting organizations that can develop and promote responsible sharing of research data. Public access to data presents a great benefit to users who have the resources and knowledge to exploit such data, either to further advance the research or to develop commercially viable products. As such, public access opens the digital archive to a global market, with no special advantage to American researchers or businesses. US job creators may therefore share a resource that can serve an expanding economy, but policy makers should recognize that the benefits cannot be confined to the US scientific community.

(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?

Data associated with scholarly research represent a class of intellectual property that can and should be distinguished from peer-reviewed research articles. Data represent information created in the course of research and subject to researchers' interpretation in the reporting of results. Traditionally, authors have transferred copyright in the peer-reviewed and copy-edited articles that report results in the publishers' chosen formats; publishers have protected copyright in these articles of record. Generally, research data have resided with the authors, who may share the data freely, or with conditions. Open access publishing models frequently recognize the author's right to retain copyright in research articles, while inviting users to freely reuse data. However, most authors will also expect attribution and credit for their primary research, including the underlying data.

(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?

Somewhat similarly to the life cycle of research articles in various disciplines, data in certain fields may have more compelling near-term value than data in other fields. This is true, for instance, in biomedical fields compared to other types of technical study. It may be reasonable to assume that the demand for publicly accessible data would parallel that of published research.

(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?

Public access policies must recognize that the archiving and curation of data carry costs, just as publishing of research does. Consistent with the Principles of Scholarly Publishing, identified above, the IEEE believes that public access models should be sustainable, whether by assured government funding, or by opportunities for repositories to charge for value-added services to offset the cost of curating publicly accessible data sets. Government funding is appropriate to creating infrastructure and propagating data interchange standards, but offsetting revenue models are desirable as a way to avoid overreliance on government support.

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

Institutions with the appropriate resources may offer services to researchers and authors to curate data sets on their behalf and for the good of scientific communities at large. Standards for interoperability among content repositories will ensure discoverability of data.

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

Recognizing that ongoing funding is needed to support hosting and maintenance of data sets, as well as assuring their availability in future formats, research grants should provide for some level of support for data archiving.

(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?

The National Science Foundation in 2011 introduced a requirement for grant recipients to file a plan for sharing of data created as a result of NSF funding. Such a plan can readily be checked by the agency for compliance. We suggest that this approach provides a sufficient process for grantees to describe their measures in a way that can be verified; allowing for grant support to promote compliance would assure participation by researchers.

(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?

As with public access to scholarly publications, care should be taken to see that small and medium-sized businesses are able to benefit from public access to funded research data. However, also in parallel with access to publications, public access to data should be sustainable through a business model that assures continued revenue to support costs of maintaining and curating a data collection. One means of support could be continuing US Government grants to support content repositories; however, government funding is less than ideal for sustainability. Encouraging private sector investment in fee-based services offered along with free access to data repositories could generate revenue that can be channeled to support the repositories themselves.

(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?

If appropriate models are adapted from traditional forms of scholarly publishing, it should be possible to achieve at least the same level of attribution and credit in reporting secondary results, but this can be validated only with experience. Attribution and provenance are critical to participation. The issue of ambiguity is an obstacle. The IEEE and other scholarly publishers are actively pursuing ways to disambiguate author names; such projects, along with ORCID (see response to Q. 11, below) will help ensure that the researcher is identified with their data.

Standards for Interoperability, Re-Use 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.

The global organization DataCite has already been successful in registering Digital Object Identifiers (DOIs) for data sets. The same continued industry collaboration that has successfully introduced the use of DOIs for articles should be brought to bear for data sets. Public access is furthered by the use of metadata that identifies US Government funding sources for research data that can be made publicly available from any content repository.

(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?

CrossRef, a not-for-profit association founded by scholarly publishers in 2002, has pioneered the development of the Digital Object Identifier (DOI) as a unique identifier to locate every published scholarly work. IEEE and other scholarly publishers would support an initiative to enhance article metadata with information clearly identifying the agency responsible for funding the described research. Agencies would save considerable effort and expense by supporting improvements to DOI metadata that would automatically capture this essential information.

Similar collaboration between public and private sectors will lead to success in efforts to provide identifiers for data sets that are created in connection with funded research. Examples include DataCite (www.datacite.org) and the NISO/NFAIS Working Group on Supplementary Journal Information (www.niso.org).

Another example of a collaborative approach among publishing partners is the Open Researcher & Contributor ID (ORCID) project (www.orcid.org), a successful public-private partnership with 275 participating organizations. This project addresses name ambiguity among individual authors and the resulting difficulties in consistent author attribution.

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

The 2009 report of the Interagency Working Group on Digital Data of the NSTC has already proposed that funding agencies require data management plans for funded projects. While recognizing that some agencies, such as the National Institutes of Health, may already have detailed policies in place, and that each agency will have unique considerations, care should be taken to approach data management in a uniform and consistent way across the spectrum of Federal agencies.

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

As with public access to scholarly publications, interoperability among content repositories would be essential to successful public access to digital data. Standards-making for data collections may be informed by the processes used to achieve interoperability for content repositories of published articles, and indeed, the same stakeholders are likely to have similar interests in the parallel cases. Government sponsorship of public-private working groups and task forces is an appropriate and productive way to bring parties together for sustained collaboration. Establishing the appropriate identifier for the publication, person and data will help facilitate the link. CrossRef is a great example of the impact the DOI has had on discovery and access.

In conclusion, IEEE appreciates the opportunity to provide input to the Working Group's deliberations and stands ready to answer questions and provide additional information as needed. If we can be of any further assistance, please contact Kenneth Moore, Director, IEEE Book & Information Services (e-mail: k.moore@ieee.org; tel: 732-562-3954).