

January 9, 2012

U.S. Office of Science and Technology Policy
Request for Information: Public Access to Digital Data Resulting From Federally Funded
Scientific Research
Docket number OSTP-2011-0022
digitaldata@ostp.gov

Office of Science and Technology Policy,

The following comments are in response to the December 23, 2011 Federal Register notification (Vol. 76, No. 247, p. 80417-80418) inviting public comment on the “Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research”.

Preservation, Discoverability, and Access

Question 1 comment

An important aspect is whether the digital data is from a specific research project has value or if there is only value from the massive aggregation of such digital data. In the case of the digital data from a specific research project then policies can be established to encourage authors to make use of a journal’s capability to make available supplementary data available for public access. In the case where only massive aggregation of such digital data has value then there need to be federal policies that encourage such massive aggregation as well as providing the resources to host such massive data sets.

Question 3 comment

Work with representative professional scientific organizations to find the most appropriate venue and manner to manage generated data. For those professional scientific organizations that also manage publications (e.g. American Chemical Society) they may be able to provide the capability to manage supporting information associated with publications.

Question 4 comment

If provided an efficient manner of making data available to their peers I believe most in the scientific community would see the value of doing so where it makes sense. I believe in the long term the scientific community will appropriately recognize those who make the effort to appropriately share their scientific data.

Question 5 comment

I believe that professional scientific organizations can best contribute through the encouragement of their members to advance the scientific enterprise by wider sharing of useful data with their fellow colleagues. Where those organizations also manage publications they can provide the capability to host associated supplementary information.

In relation to question 7, those professional scientific organizations that manage publications could provide reporting on the extent to which authors associated with specific institutions make use of the ability to provide associated supplementary information.

Question 6 comment

I believe that federal agencies, individually or collectively, should consider providing needed resources to voluntary consensus standards organizations to develop the needed digital data format standards that would greatly assist the free exchange of digital data. Separately, federal agencies would need to provide resources to provide hosting of some larger collections of digital data. While there are good examples of collections of digital data arranged on informal bases such collections are vulnerable to frequent moves and varying levels of support and maintenance.

For example the Mössbauer Effect Data Center is an example of such an informal collection of digital data. That data center is completing a move to the Dalian Institute of Chemical Physics, Chinese Academy of Science. Formerly it was located at the University of North Carolina at Asheville.

Through the provision of cloud computing resource the hosting and management of such digital data collections should be achievable on a more cost effective basis than is currently possible.

Question 7 comment

Work with professional scientific organizations that manage publications to provide reporting on the extent to which authors associated with specific institutions make use of the ability to provide associated supplementary information.

Question 8 comment

There may need to be specific funding mechanisms for federal agencies such as NSF to catalyze and initiate such endeavors.

Question 9 comment

Work with professional scientific organizations to establish policies regarding such attribution and credit policies.

Standards for Interoperability, Re-Use and Re-Purposing

Question 10 comment

The establishment of data format standards, where currently not existing or insufficient, would be needed to take maximum advantage of re-use and re-purposing of data. I believe that voluntary consensus standards organizations with appropriate assistance of federal agencies and professional scientific organizations are the best organizations to lead and maintain such developments.

Question 11 comment

I believe the development of ANSI/IEEE N42.42, “American National Standard Data Format Standard for Radiation Detectors Used for Homeland Security”, is a good example of where there was strong federal support of the standards development process and strong interactions

with commercial vendors. On the down side the focus on U.S. Homeland Security usage and limited participation by the research community has not brought that format into more general usage in the scientific community. However, the ground work has been laid and with the right additional support by federal agency, voluntary consensus standards organization, and professional scientific organizations that format could become more generally used.

For other types of radiation detection instrumentation there is a need for standard data formats to be established before there can be the free exchange of digital data that could catalyze additional scientific advances.

Question 12 comment

Federal agencies should work collectively with voluntary consensus standards organizations (e.g. ASTM International) to determine what digital data format standards are needed and support their development. Such a step would be consistent with the National Technology Transfer and Advancement Act (NTTAA), Public Law 104-113. There could also be the involvement of major professional scientific organizations (e.g. American Chemical Society). Voluntary consensus standard organizations such as ASTM International have good experience in involving participation not only from U.S. interests but also international interests.

Question 13 comment

Many scholarly publications are using DOI (digital object identifier) to uniquely allow access to specific publications. The DOI may also find utility in linking a specific publication to the associated digital data. Other linking information could be achieved through contract / grant identifiers.

For digital data that is hosted on federal agency systems there should be consideration to using a form of DOI to allow for collections of data to be easily referenced in publications. Within individual collections of digital data individual digital data elements would also need a form of DOI to uniquely identify each data element. Data collections and individual data elements should also have the ability to be versioned in those isolated instances where corrections need to be made by the generator.

Sincerely yours

Mr. Donovan Porterfield
Los Alamos, NM 87544