

Hello Ted Wackler,

I am writing to the OSTP office concerning the “Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research” that is available at <http://www.federalregister.gov/articles/2011/11/04/2011-28621/request-for-information-public-access-to-digital-data-resulting-from-federally-funded-scientific>

I will put in my comments after the numbered sections below.

## **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?

I would like to see of PubMed Central (<http://www.ncbi.nlm.nih.gov/pmc/>) to include more data as well as journal articles. With the new NSF data management plan requirements, research done with NSF funds could copy the data to an NSF repository. I would also like to see expanded roles for NTIS and the DOE Information Bridge in holding more data from research. I know that NTIS often sells their reports, but it would be better if the reports and data were freely available to the general public. Astronomical data could be held at the NASA ADS with greater Federal support, <http://adsabs.harvard.edu/index.html>

(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?

Where applicable, I would recommend that Federally funded research license their material with a CC by license (<http://creativecommons.org/licenses/by/3.0/>) or CC0 (<http://creativecommons.org/publicdomain/zero/1.0/>). This will provide the widest reach to readers throughout the whole world. This will also have the most benefit for scientists, federal agencies, the readers and the citizens of the United States. It may not be as beneficial for commercial publishers, but they have plenty of other non-government sponsored material they can publish.

(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?

There are many different data types. The Global Change Master Directory provides recommendations to scientists who deposit data to the directory. They provide guides to their metadata writers (Directory Interchange Format (DIF) Writer's Guide). See <http://gcmd.nasa.gov/User/difguide/WRITEADIF.pdf> and <http://gcmd.nasa.gov/User/difguide/difman.html>. This guide could be used as a template to help data management writers describe datasets in other disciplines.

The Digital Curation Centre is another good resource to consult, <http://www.dcc.ac.uk/resources/data-management-plans>. This is another good resource, “National initiatives for promoting data management strategies: an overview,” <http://sonexworkgroup.blogspot.com/2011/04/national-initiatives-for-promoting-data.html>

(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?

It depends on who needs to use that data, and the intended audience of the research.

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

There are many librarians who are getting to be a lot more familiar with data management plans and e-science. I would recommend that the government work with university programs such as those listed at <http://www.arl.org/rtl/eresearch/escien/nsf/nsfresources.shtml>.

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

I am not sure.

(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?

Scientists need positive reinforcement for depositing and describing their data. If they received more grant funding for cooperating in projects, or if they received greater recognition by university administrators, then that would be some positive rewards for compliance.

(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?

There are always more mashups that could be done with GIS data and social science data.

(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?

Data sets could be given a permanent citation link, such as a DOI. <http://www.doi.org/> I would recommend that you read some of the papers presented at this conference, [http://sites.nationalacademies.org/PGA/brdi/PGA\\_064019](http://sites.nationalacademies.org/PGA/brdi/PGA_064019) “Developing Data Attribution and Citation Practices and Standards: An International Symposium and Workshop”

## **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.

This chapter might be of use to you. <http://www.ncbi.nlm.nih.gov/books/NBK45678/> “The Current State of Data Integration in Science” found in the book, Steps Toward Large-Scale Data Integration in the Sciences: Summary of a Workshop. National Research Council (US) Committee on Applied and Theoretical Statistics.  
[http://www.nap.edu/catalog.php?record\\_id=12916](http://www.nap.edu/catalog.php?record_id=12916)

(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?

I can't find any right now.

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

Start with one country, and then start working with other countries. I'd recommend that you take a look at the policies of the United Kingdom. Consider looking at <http://www.dcc.ac.uk/resources/policy-and-legal/policy-tools-and-guidance> and <http://www.jiscdigitalmedia.ac.uk/crossmedia/advice/establishing-a-digital-preservation-policy/>.

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

I would recommend that you take a look at this article, <http://www.plosone.org/article/info:doi/10.1371/journal.pone.0021101> for some practices that are used.

Response to this RFI is voluntary. Responders are free to address any or all the above items, as well as provide additional information that they think is relevant to developing policies consistent with increased preservation and dissemination of broadly useful digital data resulting from federally funded research. Please note that the Government will not pay for response preparation or for the use of any information contained in the response.

Name/Email - Joseph R. Kraus

Affiliation/Organization – Science & Engineering Librarian, University of Denver, Penrose Library, 2150 E. Evans Ave.

City, State – Denver, CO 80208

In addition, please identify any other items the Working Group might consider for Federal policies related to public access to peer-reviewed scholarly publications resulting from federally supported research.

Please attach any documents that support your comments to the questions.