

Thu 1/12/2012 10:24 AM

Request for Information: Public Access to Digital Data Resulting From Federally Funded Scientific Research

Name/Email

Ahmet Erdemir

Affiliation/Organization

Computational Biomodeling (CoBi) Core

Department of Biomedical Engineering

Lerner Research Institute

Cleveland Clinic

City, State

Cleveland, OH

Comment 1

Relevant to

(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?

(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?

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

Data from federally funded research should be publicly accessible for use and for further dissemination. Ideally if one goes to a federal agency's research reporting tools, e.g. NIH RePORTER, <http://projectreporter.nih.gov/reporter.cfm>, a search should not only provide project details but also links to data/models generated by the project. For example, one can easily access list of publications generated by the project and if uploaded in PubMed Central, the publication itself. Why not have a link to data that is used in that publication? Or, links to other data that is not necessarily associated with a publication but still a result of that project? There are many federal or institutional resources to upload data, where longevity is more warranted than an investigator initiated website. For example, in our computational biomedical research we utilize SimTk (<https://simtk.org>), which is provided by the National Center for Biomedical Computing at Stanford.

Comment 2

Relevant to

(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?

While many federal agencies has initiated various data and resource sharing requirements, these does not seem to be enforced adequately. The way and the timing of the data dissemination is still at the discretion of the funded investigator. As a tax payer and a fellow investigator, I feel that when I read about the results of a federally funded publication, I should be able to contact the authors to access the

data as well. If authors cannot provide me the data, they should provide me their data/resource sharing plan (which is part of NIH submissions) that explicitly dictates why they cannot share data.

Comment 3

Relevant to

(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?

(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?

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

It is certainly possible that some investigators may not want to share data due to potential benefits from intellectual property. Nonetheless, during the submission of a federal grant, this intention should explicitly be provided and a cost sharing plan to accommodate lack of public sharing to data should be outlined. When full public access of data is planned and executed, cost sharing should not be necessary.

Comment 4

Relevant to

(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?

Investigators should be rewarded for their efforts to provide their data in a usable form. These efforts take time and resources.

Investigators are hard pressed to spend their time to write grants and to write peer-reviewed articles. The former is dictated by the desire to keep a sustaining research program and the latter is the traditional way for knowledge dissemination and the way investigators prove themselves to their institutes that they are productive. Sharing data should have the same priority as these. A while back, to accommodate this cultural thinking, we discussed on the possibility to establish a Journal for Dissemination, (see, http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Journal_for_Dissemination). We have also addressed how the dissemination can be rated. In addition, we provided a tabularized comparison of Data/Model/Software dissemination methods.

Comment 5

Relevant to

(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?

(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?

Time to dissemination and extent of data sharing is very likely to be different between disciplines; nonetheless, the principal goal should remain the same; data collected through federal funds, unless justified rigorously and supported by cost sharing, should be publicly accessible.

A relevant discipline related problem has been summarized for the patenting system and prospective reform of it (see Schacht, 2007, Patent

Reform: Issues in the Biomedical and Software Industries, Congressional Research Service Report RL33367). For example, computational sciences evolve very rapidly. A long lasting embargo to access data/models/software will likely hinder innovations by upcoming generations to innovate.

Comment 6

Relevant to

(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?

Another concept to consider in federally funded research is open development. Commonly accepted norm of doing science in the US is that, you write a grant, you do the work in your lab, you publish, and if you are generous you disseminate your data/models/software afterwards. Why not investigators, when writing a grant, are encouraged to write a plan for open science and have plans for including other stakeholders to help decision making during the research process. Open development practice by definition will dictate transparency and require early and frequent dissemination as well.

Comment 7

Relevant to

(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?

(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.

(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? Show citation box

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

The National Science and Technology Council's Interagency Working Group on Digital Data are encouraged to browse through multiscale modeling community's discussions on model and data sharing at [http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Model Sharing Working Group](http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Model_Sharing_Working_Group), [http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Data Sharing Working Group](http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Data_Sharing_Working_Group), [http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Working Group 10](http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Working_Group_10)

Comment 8

Relevant to

(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.

(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?

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

Standards are necessary for seamless exchange of data. Nonetheless, the priorities to accomplish public data access should be: 1) Share the data (so that it can be accessed), 2) Document the data (so that it can be used), 3) Conform to data standards (so that it can be exchanged easily). From an investigators perspective, as one goes from item 1 to item 3, the workload increases heavily. Along those lines, and to address item 2 for our specific discipline, we have provided reporting considerations for complicated biomechanical models (see

[http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Reporting in FEA/JB Edition](http://www.imagwiki.nibib.nih.gov/mediawiki/index.php?title=Reporting_in_FEA/JB_Edition))

Comment 9

Relevant to

(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?

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

Publication system is archaic and slow and is not always public access. The way to communicate scientific results need to adapt the speed of our evolution of social communication networks. This will require the review system and publication companies to adapt new approaches. It may be interesting to explore why a system like Wikipedia

<http://www.wikipedia.org> should not work for scientific communication.

Comment 10

Relevant to

(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?

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

Search for data should be easy. If Google or any other search engines can find the data, so does any interested parties. Innovations for filtering search results from vast variety of data sources are necessities.

Comment 11

Relevant to

(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 seems to be a wide range of innovation and business opportunities to establish federal and institutional repositories, to assemble relational databases among them, and to provide tools that will facilitate the investigators to annotate and publish their data (not just the knowledge acquired from the data).

Comment 12

Relevant to

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

A scholarly publication only provides a knowledgebase. Reproducibility, reusability, and therefore accountability are not warranted with the lack of underlying data. Some disciplines has established requirements for uploading of data with a publication. This good practice can be generalized quickly if federal agency policies prescribe provision of data with the publication.

=====

Please consider the environment before printing this e-mail

Cleveland Clinic is ranked one of the top hospitals in America by U.S.News & World Report (2010). Visit us online at <http://www.clevelandclinic.org> for a complete listing of our services, staff and locations.

Confidentiality Note: This message is intended for use only by the individual or entity to which it is addressed and may contain information that is privileged, confidential, and exempt from disclosure under applicable law. If the reader of this message is not the intended recipient or the employee or agent responsible for delivering the message to the intended recipient, you are hereby notified that any dissemination, distribution or copying of this communication is strictly prohibited. If you have received this communication in error, please contact the sender immediately and destroy the material in its entirety, whether electronic or hard copy. Thank you.