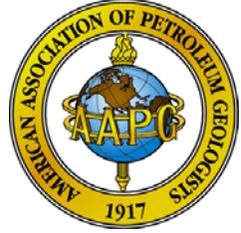


AMERICAN ASSOCIATION OF PETROLEUM GEOLOGISTS

An International Geological Organization



Dr. Paul Weimer
President

University of Colorado
Department of Geological Sciences-CB399
2200 Colorado Avenue, Benson Building
Boulder, Colorado 80309-0399
Phone: (303) 492-3809
Fax: (303) 492-2606
E-mail: paul.weimer@colorado.edu
www.aapg.org

January 12, 2012

To: Office of Science and Technology Policy
digitaldata@ostp.gov

1. The following comments are submitted by the American Association of Petroleum Geologists, a not-for profit scientific society whose mission is to advance the science of petroleum geology. Publishing is one means of fulfilling that mission. Our publications are based on research stemming from a combination of industry, and in some cases government supported research – such as NSF, USGS, and DOE.

2. As a publisher AAPG is not in the business of running a digital research data repository, and is not viewed as such by authors. AAPG as a publisher offers an online “Data Sharing Option” for our journal, the *AAPG Bulletin*. This allows authors, at their discretion, to include backup digital data which may be pertinent to their original research or to the published paper, to be posted at an online location with a linkage back to the published paper. Very few of our authors take advantage of this option.

3. The mission of many publishers, and of AAPG, is to advance science through publishing, and other means, but not in providing a repository of raw research data. As a publisher, AAPG never requires and only occasionally receives access to any of the digital data, Policies which are developed to insure the preservation and discoverability of digital data should recognize that a researcher’s raw digital data is entirely separate from the publishers publication process.

Comments to specific questions in the RFI are outlined below with the specific question outlined in **Bold**.

(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. For AAPG, and I believe most publishers, the intellectual property retained by us as a publisher rests with the publication itself, not the raw data which may have served to develop the publication.

Research data may be a combination of government, private, or industry sourced data. Policies which are developed to insure the preservation and discoverability of digital data should recognize that this effort is entirely separate from the publication process. Privately funded data, which may also be linked to the publication, would need to be excluded from the requirement of openly accessible archiving.

(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?

5. Policies for digital archival should include a definition of digital data to be provided. For example the access policy should outline whether: all raw gathered data; data which has been digitally processed or enhanced to some degree; only final tabulations of results, such as spreadsheets are to be archived. In the field of geology, for example, geochemical or other analyses may be conducted on rock samples. Several levels of analysis might be conducted on the raw geochemical measurements, ultimately leading to a tabulation of analyses, prior to interpretation and publication. Sample retention would likewise be a question. In short a policy should outline whether all or just some of these levels of data need to be retained.

6. Digital data within each scientific discipline, not to mention, across disciplines, are of uniquely different formats and standards. Data silos would need to be created within a repository, with each holding a distinct type and format of data. For example, within the field of geology, the data formats of raw geochemical analysis are vastly different from those of seismic measurements, or well log information, or analysis of a satellite image. These of course are vastly different from data sets in the fields of medicine, physics, or engineering. Perhaps a solution might be to require maintenance of only tabulated summary data in spreadsheet format, as an example; this implies no raw data but rather final processed information of some type.

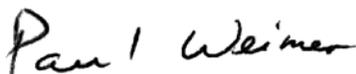
(7) What approaches could agencies take to measure, verify, and improve compliance with Federal data stewardship and access policies for scientific research? How can the burden of compliance and verification be minimized?

7. Policies for compliance of data stewardship should be outlined at the point of grant award, and to the individuals who have received that grant, and not to publishers or other third parties not involved in the research. This implies that a government coordinated or approved site would be available for the researchers to archive the digital data prior to or perhaps simultaneously with the publication process.

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

8. Scientific publishers widely utilize a DOI number (digital object identifier), which uniquely identifies that particular paper. According to the International DOI Foundation, over 55 million DOI names have been assigned by the DOI System since its inception, (<http://www.doi.org>). Government grants also have unique grant numbers, so this linkage would seem reasonable. A published paper with one DOI may tie into multiple government grants. Likewise one grant may result in multiple DOIs. Policies designed to link publications to data should focus on DOIs.

Sincerely,



Dr. Paul Weimer
AAPG President