January 5, 2012

To: Office for Science and Technology Policy (OSTP)
        digitaldata@ostp.gov

From: Gary R. VandenBos, PhD, Publisher
        gary@apa.org

        Steven J. Breckler, PhD, Executive Director, Science Directorate
        sbreckler@apa.org

        American Psychological Association
        750 First St., NE
        Washington, DC 20002

Re: **Request For Information: Public Access to Digital Data Resulting from Federally Funded Scientific Research (FR Doc. 2011-28628)**

Dear Sir or Madam,

On behalf of the American Psychological Association, we are writing to respond to the Request for Information requested by the Office of Science and Technology Policy (OSTP) in the Federal Register (Volume 76, Issue 214) of November 4, 2011, seeking public input on "approaches for ensuring long-term stewardship and encouraging broad public access to unclassified digital data that result from federally funded scientific research." We welcome this opportunity to provide recommendations regarding this matter.

As a professional association, we view public access to digital data as a primary objective. Thus, our publishing policies are moving in a direction that promotes the culture of data sharing in the field of psychology.

As a scholarly publisher, we believe that better discoverability and re-use of original research data are to be encouraged at all levels and among all stakeholders involved. As most publishers do, APA supports the view that Federal agencies should work with researchers and other stakeholders to create appropriate policies to make digital data resulting from federally funded scientific research freely available to the public. Every stakeholder has an important role to play. Governmental and other funding agencies have a special contribution to make in identifying international standards and best practices for the management of primary scientific data generated by taxpayer or other research grant funding. This role could also include standards for the interoperability of data repositories with the published research literature. To ensure that deposited datasets become an integral part of the record of science over the long term, publishers would encourage the establishment of common practices around the bi-directional linking of data and publications and around standards for the citation of data. We encourage agencies to investigate and establish contacts with a number of initiatives already underway or recently concluded that are examining data stewardship and public access issues in this context.

Along with other scholarly and professional publishers, APA recommends that Federal grants allocate specific funds to support researcher data management and deposit efforts, and to support the

establishment of discipline-specific data archives, particularly in areas such as psychology for which such mechanisms are not well developed.

Federal policies should establish clear rules for citation of data sets and acknowledgement of modifications to source data and should recognize the costs associated with hosting, maintaining and preserving raw data or data sets, and continuing to make such data available over the long term. This is important for creating incentives and rewards for data sharing, and to support recognition of data creators and collectors.

To foster greater legal certainty for data users and producers, APA recommends that Federal policy give clear direction as to what data may be shared publicly and establish penalties, e.g. grant bans for those who willfully misrepresent or distort data created by others, for the misuse or abuse of data.

Most publishers suggest that the Federal government investigate policies that create an incentives hierarchy for scientists to share their data – with the greatest reward for those who publish data with articles, but recognition also for those who publish data only or those who publish so-called descriptive data-publications

Along with other scholarly publishers, APA strongly supports the view that the Federal Government should be guided by "principles of transparency, participation and collaboration" as noted in the Transparency and Open Government Memorandum and Open Government Directive. We stand ready to work in collaboration with all partners to ensure the continued success, vibrancy, and innovation of the U.S. scientific community.

In addition to our general remarks above, we would like to comment specifically on some of the questions outlined in OSTP's Request For Information: Public Access to Digital Data Resulting from Federally Funded Scientific Research, as follows:

(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?

APA recommends that Federal policies establish clear rules for citation of data sets and acknowledgement of changes or modifications to source data. Penalties should be established for the misuse or abuse of data and technical measures that ensure ongoing data integrity should be put in place. Key policy terms should be clearly defined to differentiate between information products created for the specific display and retrieval of data ('databases') and sets or collections of raw relevant data captured in the course of research or other efforts ('data sets'). To increase legal certainty for data users and producers, clear direction should be given as to what data may be shared publicly and what may not.

Federal policies should recognize that hosting, maintaining and preserving raw data or data sets, and continuing to make such data available over the long term, has a cost which, in certain circumstances, the host should be entitled to recover. Databases themselves – i.e. collections of data specifically organized and presented, often at considerable cost, for the ease of viewing, retrieval and analysis – merit intellectual property protection, under copyright or database protection principles. These databases are often characterized by the sophistication of their data field structuring, searchability tools, and contain valuable and useful information for scholarly research. The value of individual researcher-validated data sets is different from larger-scale databases that have been organized and compiled to serve particular research needs.

(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?

APA recommends that Federal agencies should work with researchers and other stakeholders to develop appropriate policies to make digital data resulting from federally funded scientific research freely available to the public. Where no standards or commonly accepted practices for making digital data publicly available are established, APA believes that the government has an important role to play in working with key stakeholders such as researchers and publishers to develop best practices that will advance scholarly communication and the public good.

APA also recommends that those disciplines—such as psychology—that currently do not have a well-developed infrastructure for subject-specific data repositories receive financial support to establish them. To ensure that data repositories are reliable, safe and secure long term preservation of the data, such data repositories should be subject to certification and audit procedures. The facility to link datasets and publications at the level of the data set should be a condition in such certification procedures, next to all necessary preservation requirements for the long term.

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

Key elements in successfully implementing data management plans include the following:
  - o Stakeholders can engage with their communities to encourage a culture of data sharing.
  - o Data management policies established in collaboration with researchers and other stakeholders such as publishers.
  - o Requirements for data management plans should be clear, complete and unambiguous, and they should specifically address liability issues.
  - o Data management policies should consider the practices of different research communities. They should be developed in collaboration with representative bodies of all stakeholders who will likely be affected – e.g. researchers, funders, publishers, data repositories, etc. In relevant fields of science, collaboration with publishers and editorial boards can help establish clear policies for the availability of research data in the context of publications that analyze and interpret federally funded research.
  - o Training courses, e-learning modules, and FAQs should be created for researchers to gain a more complete understanding of data management plan requirements as well as the data deposit process.
  - o Specific grant funds should be available to support data management and deposit activities.
  - o Incentives for researchers to deposit data after a clearly-defined and collaboratively set time frame should be provided as well as penalties for noncompliance.
  - o Data deposit, integrity, provenance, and access at repositories should be user-friendly, efficient and clear.
  - o Bi-directional linking between datasets in data repositories and publications is to be encouraged by clear citation guidelines to ensure that datasets become part of the record of science.
  - o Data repositories should be certified and audited. Linking possibilities at the level of specific datasets should be among the certification conditions. Researchers should not be required to maintain the accuracy or integrity of the data once it has been deposited but depositing researchers should have the right to modify or correct data they have deposited. Liability policies should protect researchers if data are corrupted or lost.
  - o The administrative burden on researchers should be kept to the lowest minimum possible.

APA believes that there is no one stakeholder, (e.g. publisher, government, research institution, library, university) or data repository that has, or should have, a monopoly on any of these activities. Stakeholders should work collaboratively to address these issues.

(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?

Federal agencies must recognize the time costs associated with data attribution and credit, and provide ways to fund these efforts at both the institutional level and at the level for those receiving grants.

APA recommends the endorsement of common practice and rules for the citation of datasets and linking between datasets and publications. Federal policies should establish clear rules for citation of data sets and acknowledgement of modifications to source data. They should also provide for the establishment of security protocols that protect stored data from unauthorized modification, damage or deletion and liability arrangements if data are lost or corrupted.

In this context, APA recommends that Federal agencies seek collaborations with DataCite (see http://datacite.org/). DataCite is a non-profit organization whose aims are to establish easier access to research data on the Internet; increase acceptance of research data as legitimate, citable contributions to the scholarly record; and support data archiving that will permit results to be verified and re-purposed for future study. DataCite is currently engaged in the process of helping researchers find, identify, and cite research datasets; providing persistent identifiers for datasets, workflows and standards for data publication; and enabling research articles to be linked to the underlying data. To achieve these goals, DataCite is currently working primarily with organizations that host data, such as data centers and libraries.

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

As mentioned in the answers to other questions, APA supports the endorsement of standards for the citation of data and for the establishment of common practices around the linking between deposited datasets and related publications. In this context, APA recommends that Federal Agencies become involved with three initiatives already well underway in this area:

- o Opportunities for Data Exchange (ODE, www.ode-project.eu) – whose aim is to gather and promote best practices around the way scientific data are treated. Its Report on Integration of Data and Publications is available at http://www.alliancepermanentaccess.org/index.php/current-projects/ode/outputs/.

- o The NISO/NFAIS Supplemental Journal Materials Working Group (http://www.niso.org/workrooms/supplemental). This group is preparing an initial draft of its recommendations, which are expected to include minimum metadata elements recommended to describe supplemental materials and establish their relationship to the main article, as well as optional elements that will more comprehensively characterize materials for future applications. A non-normative Document Type Definition (DTD) is also expected in draft form. This DTD is not intended to be, or become, an official standard. Instead, it is intended for use as a model to more precisely define a hierarchy for the recommended metadata, and as a starting point for organizations seeking to adhere to the NISO/NFAIS recommendations.

- o DataCite (http://datacite.org/), a non-profit organization whose aims are to establish easier access to research data on the Internet; increase acceptance of research data as legitimate, citable contributions to the scholarly record; and support data archiving that will permit results to be

verified and re-purposed for future study. DataCite is currently engaged in the process of helping researchers find, identify, and cite research datasets; providing persistent identifiers for datasets, workflows and standards for data publication; and enabling research articles to be linked to the underlying data. To achieve these goals, DataCite is currently working primarily with organizations that host data, such as data centers and libraries.

Thank you for this opportunity to offer APA's recommendations regarding public access to digital data resulting from federally funded scientific research.