AERA Statement on Public Access to Federally Supported R&D Data

May 2013

The American Educational Research Association (AERA) is the major national scientific association of 25,000 members dedicated to advancing knowledge about education, encouraging scholarly inquiry related to education, and promoting the use of research to serve the public good. Founded in 1916, AERA as a scientific and scholarly society has long been committed to knowledge dissemination, building cumulative knowledge, and promoting data access and data sharing.

The AERA Code of Ethics mandates data sharing and acknowledgement of data use and allows for data use under restricted access provisions when necessary to protect privacy and confidentiality. Authors in AERA journals and education researchers more generally are expected to make accessible data related to their publications. They are also expected to cite data in their references to acknowledge data as scientific contributions and appropriately credit scientists.

For more than 20 years, AERA under its Grants Program funded by the National Science Foundation (NSF) has fostered the use of federally supported data sets, especially those of the U.S. Department of Education’s National Center for Education Statistics (NCES) and NSF. This long-term project has led to important scientific discoveries and methodological advances and has contributed to building scientific knowledge cumulatively through analyses of such data. Under this project, AERA now works with investigators of NSF-funded research on sharing and archiving data from completed studies on education and learning. In collaboration with the Inter-University Consortium for Political and Social Research (ICPSR), AERA is providing support and technical assistance to projects with potential for multi-investigator use and will be holding a small grants competition to stimulate use of these data. This initiative is directed to promoting data sharing and respectful, responsible use.

AERA applauds both the principles and the objectives for public access to scientific data in digital formats. We also applaud the leadership role of the Office of Science and Technology Policy (OSTP) and key science agencies like the National Science Foundation (NSF) in promoting access to data through strengthened policies and data management plans.

As emphasized almost 30 years ago by the National Research Council (NRC) in Sharing Research Data (1985), secondary analysis of extant data is essential to verification, replication, and new discoveries in science. At that time, the NRC commended various stakeholders, including federal agencies and scientific societies, to devise policies and plans for enhancing data sharing and use. One such early effort, driven in 1987 by the then NSF Division of Social and Economic Science, required grantees to commit to data sharing and archiving plans; in 1989, NSF specified a broad, agency-wide policy on data access and sharing. Since 2000 alone, the NRC has produced more than a dozen reports on expanding access to data and encouraging quality use consonant with protecting privacy and confidentiality. Also, ICPSR, now over 50 years old, has led innovations in access to useful data (including new forms of data), data preservation, appropriate use of confidential data, and data citation.

We offer our statement in this context to urge OSTP and related agencies to develop macro-level plans that not only require data management and sharing from grantees but also more broadly take steps and allocate resources to foster and facilitate a culture of data sharing and use. Knowledge about data access, the range of data amenable to sharing, and mechanisms for providing access varies within and across federal agencies and within and across fields of science. To ensure not just more policy on the books but more meaningful incorporation of policy in action requires implementation steps that can deepen and widen appreciation of the scientific value of data sharing, access, and use.
We briefly offer comments to facilitate that end. Related guidance was provided in January 2012 in response to an OSTP request for information on Public Access to Digital Data and is available at http://www.aera.net/Portals/38/docs/Publications/2AERAResponsePublicAccessDigitalDataOSTP_FR76_February2012_1-12-12_.pdf.

1. Federal policy for data sharing should include access to digital data that encompass voice and video data or other forms of big data harvested from diverse sources and preserved in digital form. Data sharing should also include the sharing of data collection instruments (e.g., interview protocols, measures, coding guides, and manuals).

2. Data Management and Sharing Plans already required by agencies like NSF are essential. Funds should be provided in awards to support archiving in data repositories to maximize data standards, access, and preservation. Renewal proposals should report on prior implementation.

3. To maximize meaningful access and contain costs, agencies should require use of data archives as the default and investigator- or institution-provided access as the exception. Agencies might offer a certified list of data archives with appropriate capacity and expertise.

4. Funds need to be provided to support repositories to expand their capacity to make accessible an expanded body of federally funded data; prepare for a larger, wider number of users; and innovate in mechanisms of access and use (as well as retiring data from use). The social sciences are fortunate to have a number of such repositories; ICPSR is the largest in holdings, innovation, and use. Fields of science with no or only limited repositories may need support to launch such entities.

5. Educational materials, webinars, or courses should be supported by science agencies, potentially in partnership with scientific societies, to provide deeper knowledge about data sharing and the value and use of third-party data archives like ICPSR. Emphasis should be placed on data sharing and principles of sound use. Also, funds for activities like the AERA/ICPSR data sharing project that enable investigators to implement data sharing plans can have high payoff and long-term impact for relatively modest cost.

6. Accessible guidance on data sharing and alignment with consent, privacy protection, and data confidentiality would be valuable. Knowledge, expertise, and views about data sharing vary widely among investigators, institutions, and institutional review boards and limit or inhibit data sharing and use. There are excellent materials from the NRC and federal statistical agencies on use of federal data bases and major federally funded data sets. An entity like the NRC might prepare a general guide for data sharing for federally funded research. The guide could also address sharing proprietary data (and working out agreements for same), big data, or administrative records when access may be affected by privacy acts (e.g., FERPA or HIPPA).

7. OSTP, federal agencies, and the Office for Human Research Protections should develop a statement to foster responsible sharing of identifiable as well as linked data as long as scientists use such data under restricted conditions, are legally bound to honor consent agreements, and face stringent penalties for disclosure. The NRC, federal agencies, data repositories, or scientific societies could assist in this task.

In conclusion, AERA urges attention to these issues and, where necessary, to the investment of cost-effective funds that can reap major scientific benefits.