



Statement on Expanded Public Access to Data

May 16, 2013

National Research Council, NAS Forum

The February 22, 2013, White House Memorandum on “Increasing Access to the Results of Federally Funded Scientific Research,” provides new opportunities for partnership between research institutions and Federal agencies. Enhanced access to digital data accelerates the pace of scientific discovery, facilitates independent confirmation of scientific results, promotes innovation, and supports education. Open data policies must also be harmonized with existing policies such as those protecting the privacy of research subjects.

Today, practices and the infrastructure for the curation, description, storage and preservation of digital data are considerably less developed in some disciplines than they are for peer-reviewed journal articles. It will be important that whatever policy or policies federal agencies propose minimizes the cost and complexity to avoid increasing the administrative overhead of compliance with grant requirements for both principal investigators and research administration. Since resources must be found within existing agency budgets to implement plans, the value of full data preservation and assured ease of access must be balanced against the costs involved, as foreseen in the first sentence of Section 4 of the Memorandum.

Coordinated federal agency policies can lead to lower barriers to digital data access, discovery, sharing and reuse. Similarly, agency policies should ensure that public access to digital data occurs through well-managed, sustained preservation archives that enable a legally and policy compliant peer-to-peer model for sharing.

While the OSTP memorandum reflects US national policy, we recognize that research is a global enterprise and the principles, tools, and solutions to accessing scholarly literature and sharing digital data will be global.

We are in agreement with OSTP that significant advantages can proceed from the open deposit of selected digital research data. However, much discussion is required to define research data before plans to build upon existing repositories, and/or create repositories to receive it, can profitably proceed.

The definition provided by OSTP in its directive is:

“ . . . data is defined, consistent with OMB circular A-110, as the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens

This is a helpful starting point for the discussion, but elaboration of this definition is required. Given the varied nature of experiments and the data gathered in different fields, and the metadata necessary to interpret correctly data gathered in a range of representations, commonly accepted standards may not yet exist in all areas covered by the directive.

We suggest that a taskforce be empanelled to give operational meaning to this definition and to address the challenges and resolve the issues OSTP enumerated in subheadings *a* to *j* of point 4 of the directive. The membership of the taskforce should include representatives from research universities, scholarly societies and federal funding agencies.

In the interim, we suggest that every federal funding agency subject to the OSTP directive develop consistent policies for researchers to include data management plans as part of their research funding proposals as soon as is feasible, so that data retention and preservation when appropriate will be universally practiced.

A data taskforce should remain on call in the future to provide advice on revising the operational definition of research data and other aspects of data management subject to the directive, since the dynamic nature of research will surely require continual updating of the definition.