

**National Science Board
Committee on Strategy and Budget
Task Force on Data Policies**

Statement of Principles

The progress of science and engineering has always been dependent on the collection of data through observation, experimentation and, more recently, computation. A core expectation of the scientific process is the documentation and sharing of results along with the underlying data and methodology, thereby allowing others to verify data, reproduce results or validate interpretations, and build upon previous work. The processes of peer review and formal publication have been pillars of scientific openness for centuries.

Recently, the increasing ease with which data can be gathered, processed, analyzed, and disseminated and funding of large-scale collaborative projects have greatly expanded the scale, scope and complexity of science and engineering data collections and highlighted the need for improved data policies. Furthermore, NSF has a commitment to broadening the participation of those involved in scientific and engineering research and education and access to data is intricately linked to this commitment. The accessibility of data created with NSF funds represents an opportunity to maximize the size and diversity of the user community for data.

The NSB is committed to the development, implementation and assessment of data sharing and data management policies for NSF-funded activities. This includes the sharing of results, data, physical collections and other supporting materials created or gathered in the course of NSF-funded work. The current policy appears in Chapter VI, Section D, of the NSF Proposal and Award Policies and Procedures Guide (pages VI-8 and VI-9 of NSF Document 10-1):

4. Dissemination and Sharing of Research Results

a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.

b. Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators. A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

c. Investigators and grantees are encouraged to share software and inventions created under the grant or otherwise make them or their products widely available and usable.

d. NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

e. NSF program management will implement these policies for dissemination and sharing of research results, in way appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like.

The Board is working with NSF leadership and other science and engineering stakeholders to frame and examine current and emerging issues associated with science and engineering data and develop relevant policies. This preliminary statement of principles will guide these efforts.

Note:

- 1. Openness and transparency are critical to continued scientific and engineering progress and to building public trust in the nation's scientific enterprise. This applies to all materials necessary for verification, replication and interpretation of results and claims, associated with scientific and engineering research.*

A strong statement about openness and transparency is an important first step.

- 2. Open Data¹ sharing is closely linked to Open Access² publishing and they should be considered in concert.*

This principle is included because there need to be bidirectional pointers between peer-reviewed and other published literature and the available supporting materials. All these materials need be made discoverable and the discoverability will require relevant metadata, ontologies, standards, etc., to be applied.

¹ Open Data refers to the concept and practice that certain data be made freely available, without restrictions, for no more than the cost of reproduction and distribution.

² Open Access publishing refers to the free availability of publications (either immediately upon publication or within a specified time period) on the public internet, permitting users to perform a variety of functions – read, download, copy, distribute, print, search, link, etc.

3. *The nation's science and engineering research enterprise consists of a broad array of stakeholders³, all of which should participate in the development and adoption of policies and guidelines.*

It is important to recognize the many different stakeholders and their respective roles and current/potential responsibilities. Their involvement in the development and implementation of policies is crucial to successful implementation.

4. *It is recognized that standards and norms vary considerably across scientific and engineering fields and such variation needs to be accommodated in the development and implementation of policies.*

The statement will be important to signal that we do not anticipate a “one size fits all” solution.

5. *Policies and guidelines are needed for open data sharing which in turn requires active data management.*

Our primary goal is the sharing of data and other supporting materials. Once available for sharing, there is a need for proactive management and preservation for long-term accessibility. The policies, roles and responsibilities vary across these different but related functions.

6. *All data and data management policies must include clear identification of roles, responsibilities and resourcing.*

These 3 R's are often omitted from consideration while the more technical aspects of policies are developed. However, in the increasingly complex scientific and engineering research enterprise, the likelihood of success will improve with consideration of the socio-economic issues that can impede or facilitate acceptance and implementation.

7. *The rights and responsibilities of investigators are recognized. Investigators should have the opportunity to analyze their data and publish their results within a reasonable time.*

³ Stakeholders include researchers, research institutions, research funders, various government agencies, professional societies, publishers, data repositories, data and metadata libraries and archives, and public advocacy groups.