



12 April 2012

Data, Information, and Librarians

Dr. Sylvia Spengler, Program Director, Division of Information and Intelligent Systems,
National Science Foundation

Data, Information, and Librarians

Sylvia Spengler
US National Science Foundation
sspengle@nsf.gov

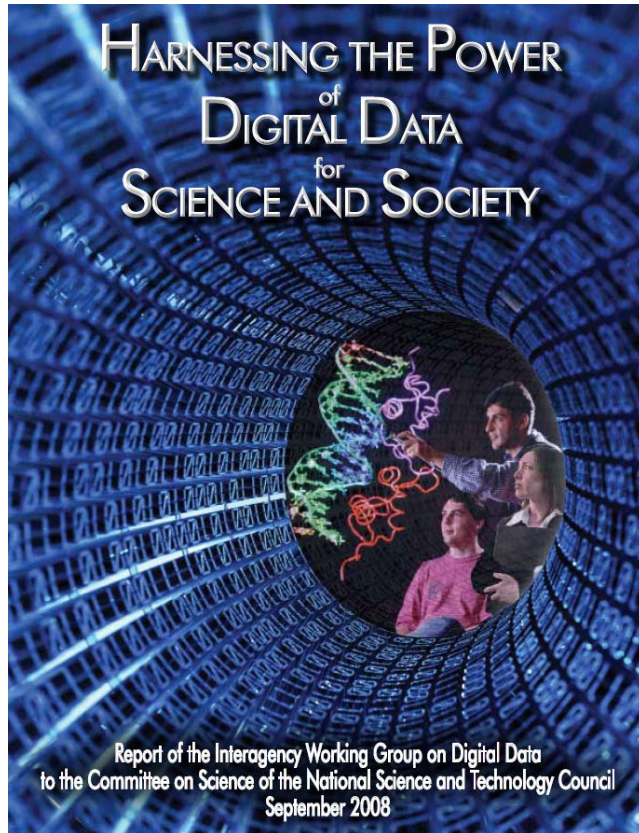


The Take-Home Messages

- US Agencies need to deal with published materials and data, linked electronically
- Libraries and librarians have extensive experience in managing and preserving physical information resources
- Electronic environments mean data need the same kinds of attention
- New skills and expertise, expanded career paths
- Data Management Plans a way to think about data at the beginning, not the end

OSTP Leadership in Public Access

- Driven by the America COMPETES Reauthorization Act of 2010
- Interagency groups on Access to Scholarly Publication AND another on Digital Data (Dept of Transportation member)
- Agencies with more than \$100 million extramural research , but other agencies play role
- RFIs on publication access and data access to results of Federally-funded research, both intramural and extramural



“The widespread availability of digital content creates opportunities for new forms of research and scholarship that are qualitatively different from traditional ways of using academic publications and research data. We call this ‘cyberscholarship’”

Guiding Principles

- Science is global and thrives in 5 dimensions
- Data are national and global assets
- Preservation is both a government and private sector responsibility and benefits society as a whole
- Communities of practice are an essential feature of the digital landscape
- Long-term preservation, access, and interoperability require full life cycle management
- Not all data need to be preserved and not all preserved data need to be kept indefinitely
- Dynamic strategies are required

Recommendations

We recommend that:

(2) appropriate departments and agencies lay the foundation for agency digital scientific data policy and make the policy publicly available

In laying appropriate policy foundations, agencies should consider all components of a comprehensive agency data policy, such as preservation and access guidelines; assignment of responsibilities; information about specialized data policies; provisions for cooperation, coordination and partnerships; and means for updates and revisions.

Why NSF Cares about Research Products

- Impact measures for Congress
- Demonstrated outcomes for the public
- Promotes Interdisciplinary science
- Cost-effective
- Promotes mechanisms for career development and contribution awareness in CV, Biosketch, etc.
- Promote different project ties through standards use, OneNSF.

How Data and Libraries Work Together

- US agencies have resources of published materials, technical reports, and data on paper
- Librarians have extensive experience with information literacy, knowledge management and curation and preservation of materials
- Librarians have extensive experience on the retrieval of information in physical publications
- Data are a type of research product, along with publications, software, reports,

- **Benefits of sharing data include:**
 - Transparency
 - Collaboration
 - Reanalysis and reuse
 - Integration or reaggregation with other data
 - Enable new model development and testing
- **Costs of sharing data include funds, time, & effort**

Long Standing NSF Policy

“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.”

(p VI-8, Award & Administration Guide, NSF 11-1)



New Implementation: Data Management Plans

- Existing policy requires awardees to share their data within a reasonable length of time & incremental cost.
- Inclusion of data management plans as a first step in what will be a more comprehensive approach to data.
- The changes are designed to address trends and needs in the modern era of data-driven science.
- NSF wants to avoid a one size fits all approach to address the issue of data sharing.

Data Management Plans

NSF requires that all proposals include a data management plan in the form of a supplemental document (maximum of 2-pages)

- Describe how the proposal will conform to NSF policy on the dissemination and sharing of research results.
- May include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification. The statement will be evaluated by peer review & program management.

New Implementation (Cont'd)

Data Management Plans

Implemented since January 18, 2011

NSF uses an automated approach in FastLane to check compliance

- Blocked submission for missing document

Data Management Plans -Oversight

Reviewed as an integral part of the proposal, coming under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.

Implementations will be reviewed in Annual Reports and Final Reports by Program Officers

Data sharing and access will become part of the Results of Prior NSF support

National Science Board calls for open access linking publications, data and software

Generic Content

The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);

Policies and provisions for re-use, re-distribution, and the production of derivatives; and

Plans for archiving data, samples, and other research products, and for preservation of access to them.

Data Management Content (2)

Policies for access and sharing Including: provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;

Requirements for content in Data Management Plans can be specific to Directorates, Offices, Divisions, Programs, or other NSF units.

NSF Policy Office has a single, searchable website that links to relevant guidance documents and examples.

<http://www.acpt.nsf.gov/bfa/dias/policy/dmp.jsp>

Data Sharing Plan – Key Elements

- **What is the purpose or goal of sharing the data?**
- **What data will be shared?**
- **Who will have access to the data?**
- **Where will the data be located?**
- **When will the data be shared?**
- **How will researchers locate and access the data?**

Data Sharing Plan – Key Elements

- **Who will have access to the data?**
 - Public at-large
 - Research community
 - More restricted (due to law or regulation)
- **Where will the data be located?**
 - Existing database
 - New database
 - Maintenance and support
- **When will the data be shared?**
 - With respect to collection
 - With respect to publication
 - Incremental release for longitudinal study
- **How will researchers locate and access the data?**
 - Availability announced via registry, publication, etc.
 - Technical protocol
 - Administrative protocol

Roles of Librarians in Electronic Information

- Education for your user community about data management, information retrieval (how-to pages, templates..)
- Consultation on how to deposit materials, may be discipline dependent
- Provision of infrastructure for publications and data from the Department (both active access and long term storage and preservation)

Current Skills and New Skills

Librarians and information scientists are content specialists (Vocabularies, other standards, metadata, descriptors, documentation)

Add a perspective on data at breadth or depth (a domain or specific context for the data and publications)

Mechanisms: continued professional development, training courses, hands-on workshops

RESOURCES

- Schools of Information or Informatics
- Library of Congress
- Institute of Museum and Library Services
 - Promotes exemplary stewardship of library collections and the use of technology to facilitate discovery
 - Advises on sustaining and increase public access to information and ideas
 - State Level programs