DATA MANAGEMENT AND SHARING

NSF's requirements for data management and sharing of the products of research:

Response by research communities served by the STS program (some highlights)

Background

- The National Science Foundation requires a data management plan (DMP) for all full proposals submitted to NSF on or after January 18, 2011.
- All full proposals must include a supplementary document of no more than two pages labeled “Data Management Plan.”
- If no data will be produced (such as a workshop proposal), the DMP should state: “No data are expected to be produced from this project.”
- The Federal government defines data in the OMB Circular A-110: “…the recorded factual material commonly accepted in the scientific community as necessary to validate research findings.” This definition includes both original data (observations, measurements etc.) as well as metadata (e.g., experimental protocols, statistical methods, etc.), and software or computer code that is required for replication, etc.

DMP Format : An Example

There is no set way for organizing DMPs. One example follows, but it would need to be tailored on a case by case basis depending on what elements are appropriate.

The DMP should (where appropriate) describe the following elements.

1. The data to be collected, and the data and metadata formats and standards to be used.

2. The physical and/or cyber resources and facilities (including third party resources) to be used to store and preserve the data after the grant ends.

3. The media and dissemination methods that will be used to make the data and metadata available to others after the grant ends.

4. The policies for data sharing and public access (including provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights as appropriate).

5. The roles and responsibilities of all parties with respect to the management of the data (including contingency plans for the departure of key personnel from the project) after the grant ends.

The Science, Technology, and Society Program (STS)

SYNOPSIS

STS supports research that uses historical, philosophical, and social scientific methods to investigate the intellectual, material, and social facets of the scientific, technological, engineering and mathematical (STEM) disciplines. It encompasses a broad spectrum of topics including interdisciplinary studies of ethics, equity, governance, and policy issues that are closely related to STEM disciplines, including medical science.

Research Communities Served by STS

There are four core communities (or societies) served by STS:

- Society for the Social Studies of Science (4S): “science" is broadly construed; includes technology, engineering, mathematics, and medicine
- History of Science Society (HSS): “science” is narrowly construed, though it does include mathematics and medicine
- Society for the History of Technology (SHOT): “technology” also includes engineering
- Philosophy of Science Association (PSA): “science” is construed as HSS

Some of the other communities served by STS are:

- Society for Philosophy of Science in Practice (SPSP)
- International Society for the History, Philosophy, and Social Studies of Biology (ISHPSSB), fondly referred to as Ishkabibble
- History of the Philosophy of Science (HOPOS)

Response to the DMP Requirement

The first STS community to respond to the DMP requirement was HSS

> See the Report on Data Management and Data-Management Plans published in the October 2013 HSS Newsletter (Vol. 42, No. 4, pp. 27-33); to locate a copy, do a Google search for “HSS Newsletter Archives”

An STS Data Management and Data Sharing Workshop was funded by the STS and held at NSF on January 29-31, 2015

> A report on the workshop, Thinking about Data Management Planning, was published on the Web on October 5, 2015

> For a copy of the report, a 4S supplemental report, and other related docs:
>
> https://cbs.asu.edu/nsf-data-management-workshop

Key Elements of the Report

The report discusses a number of important topics:

- Types of data and data management needs
- Barriers to and values of data management
- New types of possible computational research given
  - Wide adoption of data sharing practices
  - Meeting data management education and infrastructure needs
  - Promoting open access where possible
- Guidelines for structuring data management plans

Types of Data; Management Needs

- STS data take many forms ranging broadly in type from observations to interviews to sociometric analysis
- As a result, data management and sharing can have no single prescribed format
- However, it is important to strive for a general framework that satisfies both technical capacities and community standards
- Several foundational commitments can be identified
  - The PI is the individual best positioned to determine the balance between access, sensitivity, confidentiality, and the temporal needs for acquiring a complex understanding of the data
  - The field recognizes a broad time frame for the publication of data as well as its aggregated and interpreted forms
  - The field also recognizes that precisely defining “data” can be difficult (consider raw video for an ethnographic film, field notes, extended bibliographies, digital copies of un-digitized archival materials)

Barriers to Data Management

Factors that make management more difficult include:

- It takes more work and time to do data management right, or even to do it at all
- Lack of knowledge about how to manage data
- Lack of awareness of useful tools
- Lack of imagination for why this might matter and why it might help one's own work (eventually)
- Differences in willingness to share and manage data, influenced in part by subfield and career stage, and especially by availability of assistance to learn procedures and repositories and tools to maintain the data

Values of Data Management

The challenges and needed investments for establishing a meaningful data management architecture that includes individual and project specific policies and best practices as well as a shared infrastructure and standards, while substantial, are well worth the effort.

Potential payoffs include:

- data re-use
- data visualization
- computational analysis of data

Data Re-Use

Data reuse is not a common skill in STS (or any other) scientific disciplines - until we can demonstrate the value of sharing in answering recalcitrant old questions or genuinely new ones, we are unlikely to get community buy-in.

Current discussions in the context of such international bodies as the Research Data Alliance (RDA) have identified data re-use as one of the most important tasks of data science.

Data generation is getting increasingly expensive and, as new computational tools are being developed, existing data can be analyzed in novel ways, if they are preserved and curated in the right way.

That's All. Thanks You!

For a copy of these slides, please email me at:

fkronz@nsf.gov