Meeting Funders' Data Policies: Blueprint for a Research Data Management Service Group (RDMSG)

RDMSG Planning Group:
Bill Block (CISER), Eric Chen (DRSG), Jim Cordes (Astronomy), Dianne Dietrich (CUL), Dean Krafft (CUL), Stefan Kramer (CISER), Dave Lifka (CAC), Janet McCue (CUL), Gail Steinhart (CUL)

October 7, 2010

Introduction

In May of 2010, the National Science Foundation (NSF) announced its intention to require data management plans in all grant proposals (National Science Foundation 2010a). More information on the requirement was released on October 1, 2010 (National Science Foundation 2010b), with the new policy taking effect January 18, 2011. A summary of the new policy, as well as excerpts from relevant NSF documents, are provided in the appendix to this report.

The topic of sharing and managing research data has received a good deal of attention in the last few years (e.g. Association of Research Libraries 2006, Interagency Working Group on Digital Data 2009, National Science Board and National Science Foundation 2005), and with the NSF announcement, we can expect data management and sharing requirements to become the norm among major research funders. The DISCOVER Research Service Group (DRSG), in collaboration with the Center for Advanced Computing (CAC), Cornell University Library (CUL) and the Cornell Institute for Social and Economic Research (CISER), has also identified additional research data management requirements based on the specific needs of Cornell researchers. The impact of the new NSF requirement is potentially significant: Cornell researchers submit to NSF, on average, 436 proposals per year. Of these, an average of 146 proposals are funded each year with a total award amount averaging $100 million (Table 1).

Table 1. Proposal submitted to NSF and proposals funded by NSF, 2005-2009. New awards only. Data provided by Dan Dwyer.

Year	Number proposals submitted	Total proposal amount	Number proposals awarded	Total award amount (in initial year received)
2005	454	$261M	159	$146M
2006	413	$334M	132	$89M
2007	419	$330M	132	$64M
2008	434	$458M	159	$87M
2009	461	$308M	150	$113M
Average	436	$338M	146	$100M

This report summarizes the elements that we expect to be required in data management plans, describes Cornell's current capabilities and needs in meeting such requirements, and proposes a structure for a virtual organization that builds on the collaboration between the DRSG, CAC, CUL and

CISER. The proposed organization also includes Cornell Information Technologies (CIT) and Weill Cornell Medical College Information Technologies and Services (WCMC-ITS) to further develop and provide this support.

What does a data management plan look like?

NSF had not yet issued their data management plan requirement at the time this report was prepared. However, a report of the Interagency Working Group on Digital Data (IWGDD), a group of representatives from approximately two dozen federal agencies charged with developing a plan to maximize the return on investment made in federally-funded research by providing access to and preserving digital scientific data, outlined its recommendations for elements of data management plans (Interagency Working Group on Digital Data 2009). Upon reviewing NSF's new policy, we find nearly complete overlap with the IWGDD recommended data management plan elements. It is imperative to note that one of the IWGDD's guiding principles acknowledges the central importance of communities of practice, organized around their respective disciplines, in establishing best practices. The IWGDD report explicitly states that "one-size-fits-all" policies are to be avoided. The NSF policy leaves many of the details to "communities of interest," also suggesting an important place for disciplinary standards and practices. We very strongly concur with these principles, and the DRSG found this to be a widely-held view among Cornell researchers in its canvass of campus-wide data management needs.

For the purposes of this report, we use the IWGDD's data management plan elements (see box, next page) to guide our assessment of Cornell's existing capabilities and to determine additional support needed to meet the new data management plan requirement. We continue to track announcements from NSF and its various directorates, as well as other major research funders. Some NSF directorates, such as Engineering, have released their own preliminary guidelines (National Science Foundation Directorate for Engineering n.d.).

Data management plan elements recommended by the IWGDD

The following elements are very general aspects of data management that specific agencies or programs might use to guide the development of data management policies appropriate to their areas (excerpted directly from the report):

Description. Brief, high-level description of the digital scientific data to be produced.

Impact. Discussion of possible impact of the data within the immediate field, in other fields, and any broader, societal impact. Indicate how the data management plan will maximize the value of the data.

Content and Format. Statement of plans for data and metadata content and format, including description of documentation plans and rationale for selection of appropriate standards. Existing, accepted standards should be used where possible.

Protection. Statement of plans, where appropriate and necessary, for protection of privacy, confidentiality, security, intellectual property and other rights.

Access. Description of plans for providing access to data. This should include a description and rationale for any restrictions on who may access the data under what conditions and a timeline for providing access. This should also include a description of the resources and capabilities (equipment, connections, systems, expertise, etc.) needed to meet anticipated requests, including those needed for access locally, nationally and internationally.

Preservation. Description of plans for preserving data in accessible form. Plans should include a timeline proposing how long the data are to be preserved, outlining any changes in access anticipated during the preservation timeline, and documenting the resources and capabilities (e.g., equipment, connections, systems, expertise) needed to meet the preservation goals. Where data will be preserved beyond the duration of direct project funding, a description of other funding sources or institutional commitments necessary to achieve the long-term preservation and access goals should be provided.

Transfer of Responsibility. Description of plans for changes in preservation and access responsibility. Where responsibility for continuing documentation, annotation, curation, access, and preservation (or its counterparts, de-accessioning or disposal) will move from one entity or institution to another during the anticipated data life cycle, plans for managing the exchange and documentation of the necessary commitments and agreements should be provided.

Current capabilities

Initially we believe that CAC, CISER, CUL and DRSG have the capacity to provide most of the services implied by the data management plan elements listed above, and they have all participated in the development of this report. We fully expect that CIT and WCMC-ITS will also be involved as we learn more about the various types of resources that will be required to meet researchers' needs. There are

other units providing services that might support the data management needs of researchers, or advise on matters of policy (intellectual property and copyright, for example) in this area.

The following briefly summarizes some of the key services and resources for research data management plans and the units that are well suited to provide them:

Center for Advanced Computing (CAC)
- Staff expertise in the areas of data acquisition, database structure, analysis, and provenance tracking.
- Storage solutions for research data during the acquisition and analysis phases of the data lifecycle.
- Support for computing resources (private clusters, leases and pay-as-you-go) that can be used as analysis resources.
- Expertise and resources for providing secure web access to research data while leveraging web services and other standards.
- Web-based Virtual Workshop technology that can be used as an on-demand training mechanism for researchers developing research data management plans.

Cornell Institute for Social and Economic Research (CISER)
- A data archive of numeric data files in the social sciences that can accommodate Cornell research output.
- Staff expertise in metadata creation, research data organization and preservation, and analyzing, manipulating, and transforming data in the social sciences.
- Management of Cornell University's institutional membership in the Interuniversity Consortium for Political and Social Research (ICPSR). ICPSR offers researchers at member institutions the option of depositing data into its archive, which has been a trusted repository of social science datasets since the 1960s.
- Management of the Cornell Restricted Access Data Center (CRADC), a secure computing environment for use of restricted-use data.

Cornell University Library (CUL)
- Digital collections and repositories that may be used to meet data access requirements, including an institutional repository (eCommons), the Cornell University Geospatial Information Repository (CUGIR), and a data "staging" repository (DataStaR).
- Dedicated metadata services unit that provides metadata consulting and creation services.
- Staff expertise in the data management standards and practices of several disciplines.
- Responsibility for long term preservation and stewardship of the University's intellectual assets, and a nearly completed preservation repository for digital content.

DISCOVER Research Service Group (DRSG)
- A forum for identifying and helping to implement new methods for the curation, preservation, and mining of scientific data.
- Bridging gaps and integrating solutions between service providers and/or across disciplines
- Case studies via pilot projects in specific research domains that use cyberinfrastructure elements, including large-volume data sets, accessibility via high-speed network, database generation, and aggregation of data and metadata for research driven queries.

- Research and development in large-scale cyberinfrastructure.

Cornell Information Technologies (CIT)
- To be determined.

Weill Cornell Medical College – Information Technology Services (WCMC-ITS)
- To be determined.

Gap Analysis

While the multiple units provide significant expertise and services for managing research data, there are several important gaps identifiable by reviewing the list of data management plan elements (box, page 3):

Content and Format. Individuals with expertise in data and metadata content and format for various disciplines work within the service units (e.g. Gail Steinhart, Environmental Sciences & Research Data Librarian), as well as within academic departments (e.g. Adam Brazier, Research Associate in Astronomy). The gap relates to coverage, communication and coordination. Identification of individuals willing to serve as consultants in their areas of expertise is needed, as well as a means for communicating and coordinating among the network of potential consultants.

Protection. The IWGDD groups several very distinct issues of protection into this single element. In terms of intellectual property rights, the library is a logical center for issues surrounding copyright and issues related to expressing usage rights and conditions for data, and will need to designate a staff member to develop and maintain expertise in this area. The Cornell Center for Technology Enterprise and Commercialization (CCTEC) may be the logical point of contact for commercialization and licensing issues. Confidentiality and privacy are areas where Cornell may not have sufficient capacity to meet researchers' needs, although CISER and the Survey Research Institute have expertise in these areas, and the Institutional Review Board may also have a role. In addition, faculty members have areas of special expertise that can be a valuable resource for Cornell, such as John Abowd's expertise in data anonymization. Collaborations with these faculty can help ensure Cornell's competitiveness on grants.

Access. There are several solutions available on campus, but these likely do not meet the needs of all users:
- CUL's institutional repository, eCommons, for small-scale data sets in any format.
- CUL's data staging repository (DataStaR, in development) for limited-term sharing and as a means for publishing data sets to selected external repositories.
- CISER's data archive, for social science data.
- Customized solutions from CAC, CISER, CIT and CUL on a fee-for-service basis.

Publication to external repositories, when available (e.g. GenBank, National Virtual Observatory, etc.) is an additional possibility for satisfying access requirements.

Preservation. This is potentially a very significant gap. Short-term storage might reasonably be written into and covered by research grants, but storage alone is not sufficient to ensure the long-term usability of digital data. Effective long-term preservation encompasses curation and stewardship activities such as

developing and maintaining preservation plans, monitoring the integrity of stored content, monitoring and implementing appropriate technologies and standards, migrating media and file formats, reappraisal of content, and more. We still do not fully understand the full cost of digital preservation, and costs for preserving digital data will continue to be incurred well after a research grant ends. Two possible solutions, for some data sets, are CUL's preservation repository and the CISER Data Archive (in existence since 1981). What types of data will be accepted, how costs will be determined and recovered, how the curation process will be supported and managed, and how the staff and resources required to meet future requirements will scale are not yet known.

Transfer of Responsibility. Initially this is the responsibility of the PI, but there is significant potential for a "responsibility gap." Responsibility over the course of the data lifecycle must be agreed upon, budgeted for and tracked by all responsible parties when a project begins. Ensuring our responsibilities are met and can be demonstrated in future audits is an area that may require the development of new university policies.

In order to scale existing services appropriately and develop needed new ones, we need sensible estimates of the impact of NSF's new requirement. We propose that a survey of NSF awardees at Cornell be undertaken as soon as possible to obtain information to help inform planning activities. Questions should be organized around the data management plan elements described above and/or those specified in the new NSF policy. We should be able to gauge the number of affected researchers, what campus-based services researchers would require and utilize, and to estimate the quantity of data involved. Service providers will be able to use this information to estimate costs and develop or revise service models.

While the NSF-wide requirement demands our immediate attention, it also makes sense to consider the data management requirements of specific programs and directorates within NSF, as well as those of other agencies. Two graduate students from the School of Information Studies at Syracuse University are beginning work with CUL librarians Dianne Dietrich and Gail Steinhart to locate data management policies of major funders and to summarize their requirements. If time and resources permit, the analysis may be extended to include selected journals and publishers. This work will give us a more general sense of data management requirements researchers are facing.

Proposed model for a research data management services virtual organization

Our goal is to make it as simple as possible for researchers to obtain the services they require. This will first and foremost demand coordination and cooperation among service providers, and to that end we propose the creation of a new service group: a Research Data Management Service Group (RDMSG). It is imperative to present a coherent set of services to researchers, which will entail developing a unified web presence providing general information on data management planning, services available on campus, standard language that may be used in data management plans in grant proposals, and a single point of contact that puts researchers in touch with specialized assistance as the need arises. A public directory of research data support service providers was compiled by CUL staff and is already available online: https://confluence.cornell.edu/display/datasupp/Home. While this directory is a start in this direction, it is insufficient for two reasons. First, the current directory leaves it up to researchers to navigate the collection of links to put together the services they need. Second, new requirements for data management plans will place demands on service providers that were perhaps not anticipated at the time that the directory was developed. Service providers need information to understand how much

the demand for their services might increase. Scaling up may require service providers to make changes to how they manage, provide, and charge for those services, and improve or increase staff and infrastructure. Service providers will also need a chance to consider how to define and fulfill their role in the new virtual organization.

Developing the RDMSG will require coordination and cooperation among existing service providers (Figure 1). We propose the following governance and operations structure (Figure 2):

- Joint sponsorship of the new service group by the Senior Vice Provost for Research and the University Librarian.
- An RDMSG management council to coordinate and develop services.
- A faculty advisory board to ensure that the needs of researchers are met across the disciplines.
- A coordinator to facilitate the work of the management council and the day-to-day operations of the RDMSG.
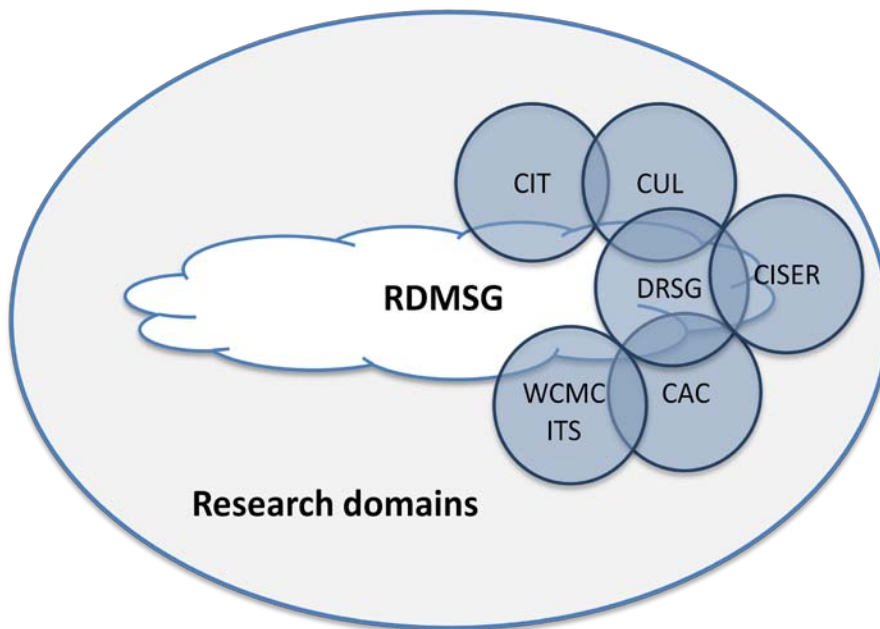

Figure 1: Diagram of the participation of the various units that can provide services to Cornell researchers. The partial overlap of the proposed RDMSG with other units signifies that those units have an interest in research data management along with other interests and responsibilities. Researchers from various domains may interact with any of the groups and services indicated in the figure, via the RDMSG, or directly.
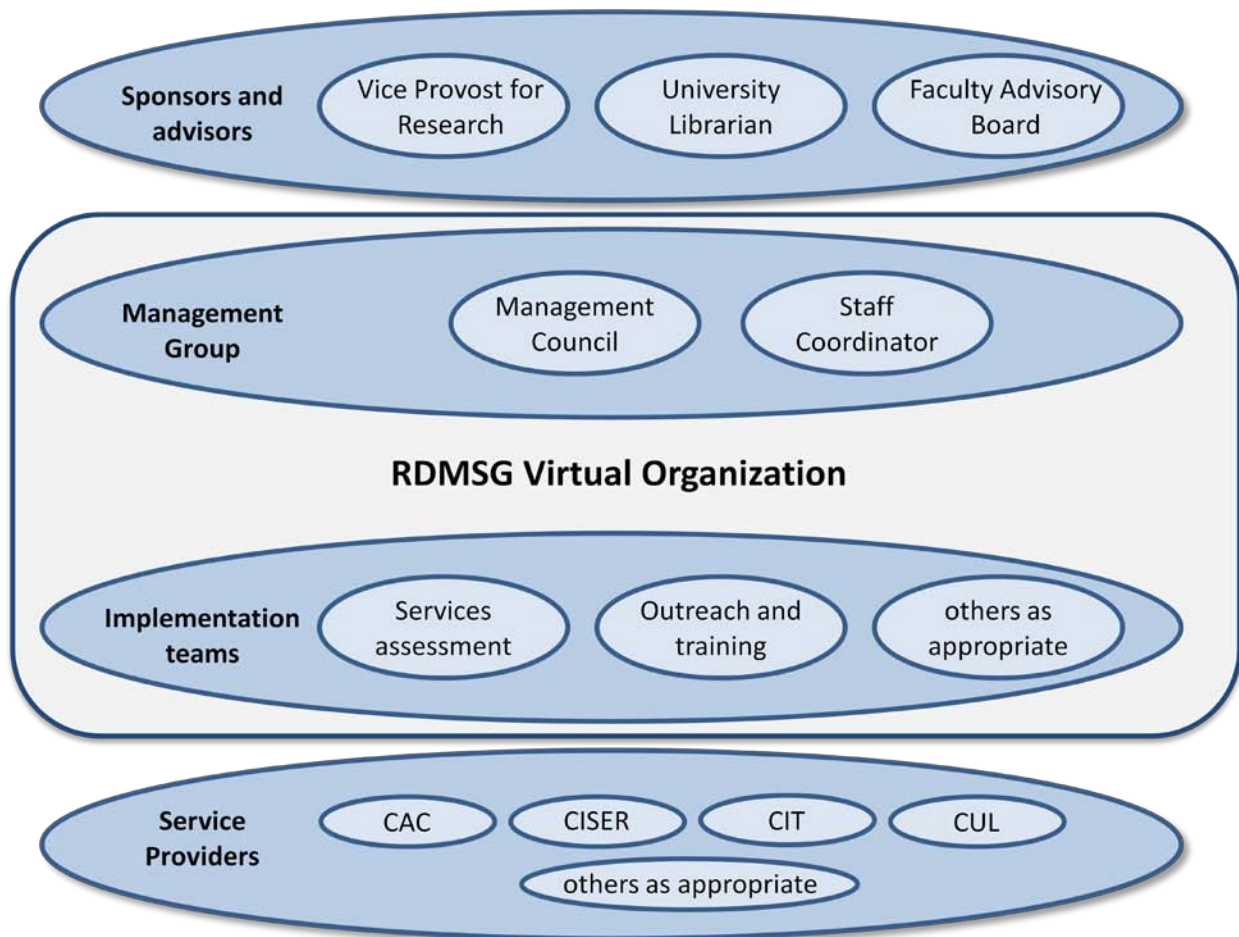
.

Figure 2. Governance and operations structure of the proposed RDMSG. The Senior Vice Provost for Research, University Librarian, and Faculty Advisory board advise the RDMSG. The RDMSG virtual organization itself is composed of the Management Council and Staff Coordinator, as well as implementation teams. Implementation teams are composed of staff drawn from cooperating service providers, and complete work or projects that directly support the work of the RDMSG. Service providers offer services as part of the RDMSG, but have additional services and activities that are not part of the RDMSG.

RDMSG management council - The council should be comprised of representatives with decision-making authority from each of the major service providers (CAC, CUL, CISER, CIT) and stakeholders (CIOs from Ithaca and Weill campuses), additional service providers as needed, and for at least the first year of the existence of the RDMSG, representatives from the DRSG. The membership of this group may change in composition and in size over time. We expect the initial council to include:

- Bill Block (Director, CISER)
- Eric Chen, RDMSG coordinator, ex-officio
- Curtis Cole (CIO, Weill Cornell Medical College)
- Jim Cordes (Astronomy, DRSG)
- Dave Lifka (Director, CAC)

8

- Dean Krafft (Chief Technology Strategist, CUL)
- Janet McCue (CUL, DRSG)
- Steve Schuster, interim (CIO, Cornell University Ithaca campus)
- Others – as the need for new service/resource providers become necessary

The management council will play a critical role in evaluating information on research activity to determine the impact of new funder requirements, developing the service group by aligning existing services with new requirements and recommending additional services as needed, developing a web portal for researchers, coordinating outreach to researchers, and tracking the progress and successes of the service group. Management Council members will charge implementation teams comprised of staff from their respective units to carry out specific tasks required to develop and support the RDMSG. Participation in the management council or on implementation teams is an inherent overhead cost to support research at Cornell, and these activities should be considered within the normal scope of responsibilities of all participants.

Specific areas of work for the management council (and the teams it creates) include:

- Conducting the survey of NSF awardees mentioned above. It is desirable to begin this effort as soon as possible, perhaps in advance of fully organizing the management council. The management council will review the results and use them to assess current capacity and plan new services.
- Updating the list of known services and service providers and assessing whether existing services are sufficient to meet anticipated needs.
- Recommend changes to or augmentation of existing services, and proposing new services as appropriate.
- Overseeing development of web presence for the RDMSG, which will host materials developed by the training and outreach group (below) and provide information on service providers and their services and fees.
- Developing and overseeing a training and outreach program for researchers. The team the council selects will be charged with tasks in this area and will create examples of data management plans, create data management planning checklists and flow charts, offer practical training and tutorials to researchers, and participate or initiate in other outreach and training activities as needed.
- Developing a process for routing researchers' requests for assistance with data management planning and services to the appropriate service providers.
- Communicating effectively with all service providers and other involved units such as OSP, IRB and OVPR.
- Developing a longer-term sustainability and financial plan to support the ongoing costs of data curation and preservation, a very significant challenge as the full cost of these activities is not known.

Faculty advisory board – A faculty advisory board will be required to ensure the RDMSG provides the necessary resources and services to support data management plan requirements. This board will interface with the management council and its sponsors and advisors (Figure 2). We recommend an initial board with at least three members, recruited by the Senior Vice Provost for Research and/or the University Librarian. Advisory board members would ideally be involved with data-intensive research,

and may already have working relationships with CAC, CISER, or CUL. We also recommend liaison or ex-officio members from the Office of Sponsored Programs and the Institutional Review Board.

RDMSG Coordinator - The RDMSG will require central coordination and support (under the direction and oversight of the management council) in order to work effectively. Initially, we propose a staff position whose key responsibilities include this central coordination role. Resources will be adjusted as needed. The necessity for continued support will be reviewed by the management council and sponsors at yearly intervals. The RDMSG coordinator will:

- Serve as the primary point of contact for the RDMSG.
- Have primary responsibility for development of the RDMSG web site.
- Work with the management council, the training and outreach team, and specific service providers to develop template language and budget guidance for data management plans.
- Maintain a directory of current contacts for each service provider in order to facilitate communication between the RDMSG and service providers (particularly service providers or other stakeholders that might not be represented on the management council), and obtain information from providers on services, fees, etc.

Reporting and accountability – Under the direction of the management council, the RDMSG staff coordinator will track outputs and successes of the RDMSG and solicit feedback from researchers to gauge performance. This will be accomplished in part by deploying tracking software so that metrics of success (e.g. number of researchers supported and number of grants submitted/won with RDMSG support) can be tracked and used to justify effort and resource expenditures over time, and to ensure that credit is attributed to both the RDMSG and participating services providers when that is appropriate.

Timeline through 2011 – The new requirement from NSF is expected to take effect in January 2011, and the average time to notification of an award is 300 days. That means the most immediate needs of researchers will be for information and assistance in developing competitive data management plans, and the real needs for data management services should begin to ramp up in Fall 2011. We recommend forming the Management Council and Faculty Advisory board as soon as possible, and that the management council form and charge teams with the most time-sensitive of tasks associated with the areas of work described above.

References

Association of Research Libraries (ARL) Workshop on New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe. 2006. To stand the test of time: Long-term stewardship of digital data sets in science and engineering. Online: http://www.arl.org/info/events/digdatarpt.pdf

Interagency Working Group on Digital Data. 2009. Harnessing the power of digital data for science and society: Report of the interagency working group on digital data to the committee on science of the national science and technology council. Washington, D.C. Online: http://www.nitrd.gov/about/Harnessing_Power_Web.pdf

National Science Board and National Science Foundation. 2005. Long-lived digital data collections. Washington, D.C. Online: http://www.nsf.gov/pubs/2005/nsb0540/start.jsp

National Science Foundation. 2010a. Scientists seeking NSF funding will soon be required to submit data management plans - US national science foundation (NSF) press release 10-077. Online: http://www.nsf.gov/news/news_summ.jsp?cntn_id=116928&org=NSF&from=news

National Science Foundation. 2010b. Dissemination and Sharing of Research Results. Online: http://www.nsf.gov/bfa/dias/policy/dmp.jsp (accessed 10/6/2010).

National Science Foundation Directorate for Engineering (ENG). n.d.
Data Management for NSF Engineering Directorate Proposals and Awards. Online: http://nsf.gov/dir/index.jsp?org=ENG

Appendix: Excerpts from NSF Documents Related to Data Management Plans

Summary and interpretation of NSF documents related to data management plans:

- The data management plan is a supplementary document not to exceed two pages in length, and will be required of all grant proposals submitted or due on or after January 18, 2011. See "Plans for data management and sharing the products of research," below.
- Agency-wide guidance for preparing a data management plan is very general. Applicants are asked to describe the types of data to be collected, standards to be used for data and metadata, policies for access and sharing (including protection of privacy), policies and provisions for reuse, and plans for archiving and preserving access to data. See "Plans for data management and sharing the products of research," below.
- Some individual directorates and programs within NSF have more specific guidelines, and we can expect to see more of this in the future.
- What is considered appropriate data management is left largely to individual disciplines, via the peer review process and program management. See "Data Management & Sharing Frequently Asked Questions (FAQs)," below.
- The primary mechanism to ensure post-award compliance appears to be the requirement that PIs and co-PIs describe the availability of research products in subsequent grant proposals. See "Results from Prior NSF Support," below.
- Costs associated with data management are allowable and should be included in the budget and budget justification. These costs will generally be appropriately specified on line G2 (publication/documentation/dissemination). See "Data Management & Sharing Frequently Asked Questions (FAQs)," below.

Plans for data management and sharing of the products of research
From: Grant Proposal Guide: Chapter 2 Proposal Preparation Instructions (II.C.2.j)
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#IIC2j

j. Special Information and Supplementary Documentation

Plans for data management and sharing of the products of research. Proposals must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see AAG Chapter VI.D.4), and may include:

1. the types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. 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);
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for re-use, re-distribution, and the production of derivatives; and
5. plans for archiving data, samples, and other research products, and for preservation of access to them.

Data management requirements and plans specific to the Directorate, Office, Division, Program, or other NSF unit, relevant to a proposal are available at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp. If guidance specific to the program is not available, then the requirements established in this section apply.

Simultaneously submitted collaborative proposals and proposals that include subawards are a single unified project and should include only one supplemental combined Data Management Plan, regardless of the number of non-lead collaborative proposals or subawards included. Fastlane will not permit submission of a proposal that is missing a Data Management Plan. Proposals for supplementary support to an existing award are not required to include a Data Management Plan.

A valid Data Management Plan may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification. Proposers who feel that the plan cannot fit within the supplement limit of two pages may use part of the 15-page Project Description for additional data management information. Proposers are advised that the Data Management Plan may not be used to circumvent the 15-page Project Description limitation. The Data Management Plan will be 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.

Results from Prior NSF Support
From: Grant Proposal Guide: Chapter 2 (II.C.2.d)
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#IIC2d

(iii) Results from Prior NSF Support
If any PI or co-PI identified on the project has received NSF funding in the past five years, information on the award(s) is required. Each PI and co-PI who has received *more than one award* (excluding amendments) must report on the award most closely related to the proposal. The following information must be provided:

(e) evidence of research products and their availability, including, but not limited to: data, publications, samples, physical collections, software, and models, as described in any Data Management Plan (…)

Reviewers will be asked to comment on the quality of the prior work described in this section of the proposal. Please note that the proposal may contain up to five pages to describe the results. Results may be summarized in fewer than five pages, which would give the balance of the 15 pages for the Project Description.

Dissemination and Sharing of Research Results
From: Award & Administration Guide Chapter VI - Other Post Award Requirements and Considerations (VI.D.4)
http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4

D. Intellectual Property
4. Dissemination and Sharing of Research Results

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.

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 ways 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.

Data Management & Sharing Frequently Asked Questions (FAQs)
http://www.nsf.gov/bfa/dias/policy/dmpfaqs.jsp

1. What constitutes "data" covered by a Data Management Plan?
What constitutes such data will be determined by the community of interest through the process of peer review and program management. This may include, but is not limited to: data, publications, samples, physical collections, software and models.

2. Is a plan for Data Management required if my project is not expected to generate data or samples?
Yes. It is acceptable to state in the Data Management Plan that the project is not anticipated to generate data or samples that require management and/or sharing. PIs should note that the statement will be subject to peer review.

3. Am I required to deposit my data in a public database?
What constitutes reasonable data management and access will be determined by the community of interest through the process of peer review and program management. In many cases, these standards already exist, but are likely to evolve as new technologies and resources become available.

4. There is no public database for my type of data. What can I do to provide data access?
Contact the cognizant NSF Program Officer for assistance in this situation.

5. Should the budget and its justification specifically address the costs of implementing the Data Management Plan?

Yes. As long as the costs are allowable in accordance with the applicable cost principles, and necessary to implement the Data Management Plan, such costs may be included (typically on Line G2) of the proposal budget, and justified in the budget justification.

6. My institution's policy is that the data and all supporting materials from all research are owned and must remain with the institution if I leave. How does this policy affect what I can say about data management and access?
Data maintenance and archiving by an institution is one avenue by which data preservation and access can be achieved. However, the data access plan must address the institutional strategy for providing access to relevant data and supporting materials.

7. Does data management and access include supporting documentation and metadata, such as validation protocols, field notebooks, etc.?
All researchers are expected to be able to explain and defend their results. Doing so usually entails maintaining complete records of how data were collected. The manner in which one maintains such records and makes them available to others will vary from project to project. What constitutes reasonable procedures will be determined by the community of interest through the process of peer review and program management. These standards are likely to evolve as new technologies and resources become available.

8. How long should data be archived and made accessible?
What constitute reasonable procedures will be determined by the community of interest through the process of peer review and program management.

9. Does this policy mean that I must make my data available immediately, even before publication?
Not necessarily. The expectation is that all data will be made available after a reasonable length of time. However, what constitutes a reasonable length of time will be determined by the community of interest through the process of peer review and program management.

10. What are NSF's expectations regarding the release of data that include sensitive information (e.g., information about individuals or locations of endangered species)?
Such data must be maintained and released in accordance with appropriate standards for protecting privacy rights and maintaining the confidentiality of respondents. Within legal constraints, what constitutes reasonable data access will be determined by the community of interest through the process of peer review and program management.

11. My data include information of potential commercial value. Am I required to make that information available?
Not necessarily. It is NSF's strong expectation that investigators will 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. However, it is also necessary to protect intellectual property rights and potential commercial value. The Data Management Plan should describe the proposed approach, which will then be subject to peer review and program management. (For example, research use of sensitive data is often allowed through reasonable binding agreements that contain confidentiality provisions.)

12. Does NSF have particular requirements for archiving and accessibility of samples, physical collections and so forth?

No. If appropriate, your Data Management Plan should describe the types of samples, and/or collections, etc., that you will use, as well as personal, institutional or other repositories for archiving and providing access to others. What constitutes reasonable archiving and accessibility will be determined by the community of interest through the process of peer review and program management.

13. Does NSF have particular requirements for what types of samples, physical collections, and so forth should be saved?
No. What constitutes reasonable requirements will be determined by the community of interest through the process of peer review and program management. These standards are likely to evolve as new technologies and resources become available.

14. If data or samples are requested before I have completed all analyses on them, must I share them?
No. The expectation is that all data will be made available after a reasonable length of time. One standard of timeliness is to make the data or samples accessible immediately after publication. However, what constitutes a reasonable length of time will be determined by the community of interest through the process of peer review and program management

15. How does this policy relate to the issue of open access publishing?
Open-access publishing (making all published articles freely available) is a separate issue that is not addressed in the implementation of the data management plan requirement.

16. If I participate in a collaborative international research project, do I need to be concerned with data management policies established by institutions outside the United States?
Yes. There may be cases where data management plans are affected by formal data protocols established by large international research consortia or set forth in formal science and technology agreements signed by the United States Government and foreign counterparts. Be sure to discuss this issue with your sponsored projects office (or equivalent) and your international research partner when first planning your collaboration.