The RDMSG: Data Management Planning and More

Wendy Kozlowski
Sarah Wright

February 2014
Support for Cornell investigators

Research Data Management Service Group (RDMSG)

- Virtual organization including CAC, CUL, CISER, IT@Cornell and others.
- Jointly sponsored by the Senior Vice Provost for Research and the University Librarian.
- Management Council and Faculty advisory board.
- Single point of contact to request assistance with data management planning.

rdmsg-help@cornell.edu
Operating Principles

- Provide **timely and professional assistance**
- Encourage **best practices** in data management
- Share information with other RDMSG consultants to provide best possible service
  - Treat information in grant proposals as confidential
- Refer researchers to the most appropriate services, at Cornell or elsewhere

rdmsg-help@cornell.edu  data.research.cornell.edu
What can we help with?

- Proposal-related (and other) data management plans
- Finding & understanding storage & backup options
- Deciphering data licensing options
- Referrals to IP & patent law experts
- Suggest data handling tools & resources
- Data description & metadata generation
- Data sharing & archival

rdmsg-help@cornell.edu  
data.research.cornell.edu
Data Sharing Not New

“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.”

“The OSTP hereby directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government.”
Publishers are suggesting or requiring data sharing

**Nature:**
- “...authors are required to make materials, data and associated protocols promptly available to readers...”

http://www.nature.com/authors/policies/availability.html

**Science:**
- “All data necessary to understand, assess, and extend the conclusions of the manuscript must be available to any reader of Science.”

http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml#dataavail
Other funding agencies with DMP-like requirements:

- Gordon and Betty Moore Foundation
- Gulf of Mexico Research Institute
- IMLS
- NIH
- NOAA
- NEH (Office of Digital Humanities)
- Smithsonian Institute Digital Asset Management Plan
- USGS
NSF’s Data Sharing Policy

- As of January 18, 2011, all proposals must include a data management plan (DMP).
- Supplementary document; 2 page limit.
- FastLane won’t accept proposals without a DMP.
- Subject to peer review (intellectual merit and/or broader impacts).

http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#IIC2j
NSF’s Data Sharing Policy

“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.”

http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp#IIC2j
NSF’s Data Sharing Policy

Specifics are left to “communities of interest” as determined by peer review and program management.

Some directorates, programs or specific RFPs may offer more specific guidance.

NSF’s Data Sharing Policy

Compliance: 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.”

“...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 ...”

http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp

rdmsg-help@cornell.edu
data.research.cornell.edu
NSF’s Data Sharing Policy

Compliance: Biographical sketches for senior personnel

“A list of: (i) up to five products most closely related to the proposed project; and (ii) up to five other significant products(...). Acceptable products must be citable and accessible including but not limited to publications, data sets, software, patents, and copyrights. Unacceptable products are unpublished documents not yet submitted for publication, invited lectures, and additional lists of products. (...)

Each product must include full citation information including (where applicable and practicable) names of all authors, date of publication or release, title, title of enclosing work such as journal or book, volume, issue, pages, website and Uniform Resource Locator (URL) or other Persistent Identifier.”

http://nsf.gov/pubs/policydocs/pappguide/nsf13001/gpg_2.jsp
Who pays:

• Costs associated with data management are allowable.
• Include on line G2 (publication, documentation, dissemination).
• Explain in budget justification.

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

Suggested practice:
• Describe the data to be produced in the course of the proposed project.
• Indicate which will be shared and at what stage (raw, processed, analyzed).
• Describe why the data are of interest to a broader community (impact).

Examples of Cornell resources:
• Best done by PI, but RDMMSG consultants can help.
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);

Suggested practice:
- Identify formats of data files created.
- Select file formats for sharing that maximize reuse and longevity; describe plans for conversion to those formats.
- Describe metadata standards to be used, who will create metadata and when.

Examples of Cornell resources:
- RDMSG consultants
- Metadata Services
- Digital Consulting & Production Services
- RDMSG guide on preparing tabular data for description and archiving
3. policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;

Suggested practice:
• NSF recognizes legal and ethical requirements may preclude sharing of some data. Trade secrets and commercial information also not subject to the data sharing requirement.
• Describe data management practices that will be used to maximize security and protect privacy, if applicable.

Examples of Cornell resources:
• RDMSG white paper on intellectual property issues
• Survey Research Institute
• RDMSG consultants
• Cornell University Policy Office
• Copyright Information Center

rdmsg-help@cornell.edu    data.research.cornell.edu
NSF’s Data Sharing Policy

4. policies and provisions for re-use, re-distribution, and the production of derivatives;

Suggested practice:
• Describe how you will make it possible for others to gain access to and use your data.
• Describe usage rights, licenses, or other policies related to re-use and redistribution of data.

Examples of Cornell resources:
• RDMSG consultants
• eCommons digital repository
• RDMSG white paper on intellectual property issues
NSF’s Data Sharing Policy

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

Suggested practice:
• Describe means by which you will provide access to data and applicable time frame.
• Describe means for preserving data, if different from above.

Examples of Cornell resources:
• RDMSG consultants
• eCommons digital repository
• IT@Cornell
• Center for Advanced Computing
• Cornell Institute for Social and Economic Research (CISER)
NSF’s Data Sharing Policy

Common questions:

• What “counts” as data?
  • Community norms
  • Reproducibility
  • Reuse

• I already publish my data in journals. Is this something different?
You can help us provide better service:

• Don’t wait until the very last minute.
• Share as much information as you can about the proposal.
• Let us know the outcome of your proposal.
• Let us know about developments related to this new requirement in your discipline, directorate or program.
• Let us know what you learn as a reviewer.
• Let us know what services you need.
• Consult with us on subsequent proposals.

Additional Resources:
• CDL’s DMPTool
• ICPSR’s Guidelines for Effective Data Management Plans
Cornell University
Research Data Management Service Group

Find us:
via email: rdmsg-help@cornell.edu
online: http://data.research.cornell.edu