



Cornell University

Institutional Support for Data Management Requirements

Gail Steinhart
Research Data & Environmental Sciences Librarian
Cornell University Library

Coalition for Academic Scientific Computation Meeting
March, 2011



Existing 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.”

http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp#VID4



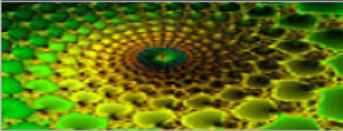
SEARCH

NSF Web Site



HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT | FastLane

**Office of Budget,
Finance and Award
Management (BFA)**



[DIAS Home](#)

[CAAR Branch](#)

[Policy Office](#)

[Systems Office](#)

[View DIAS Staff](#)

Search DIAS Staff



BFA Organization

[Office of Budget, Finance, &
Award Management](#)

[Budget Division](#)

[Division of Acquisition and
Cooperative Support](#)

[Division of Financial
Management](#)

Dissemination and Sharing of Research Results

NSF Data Sharing 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. See [Award & Administration Guide \(AAG\) Chapter VI.D.4](#).

NSF Data Management Plan Requirements

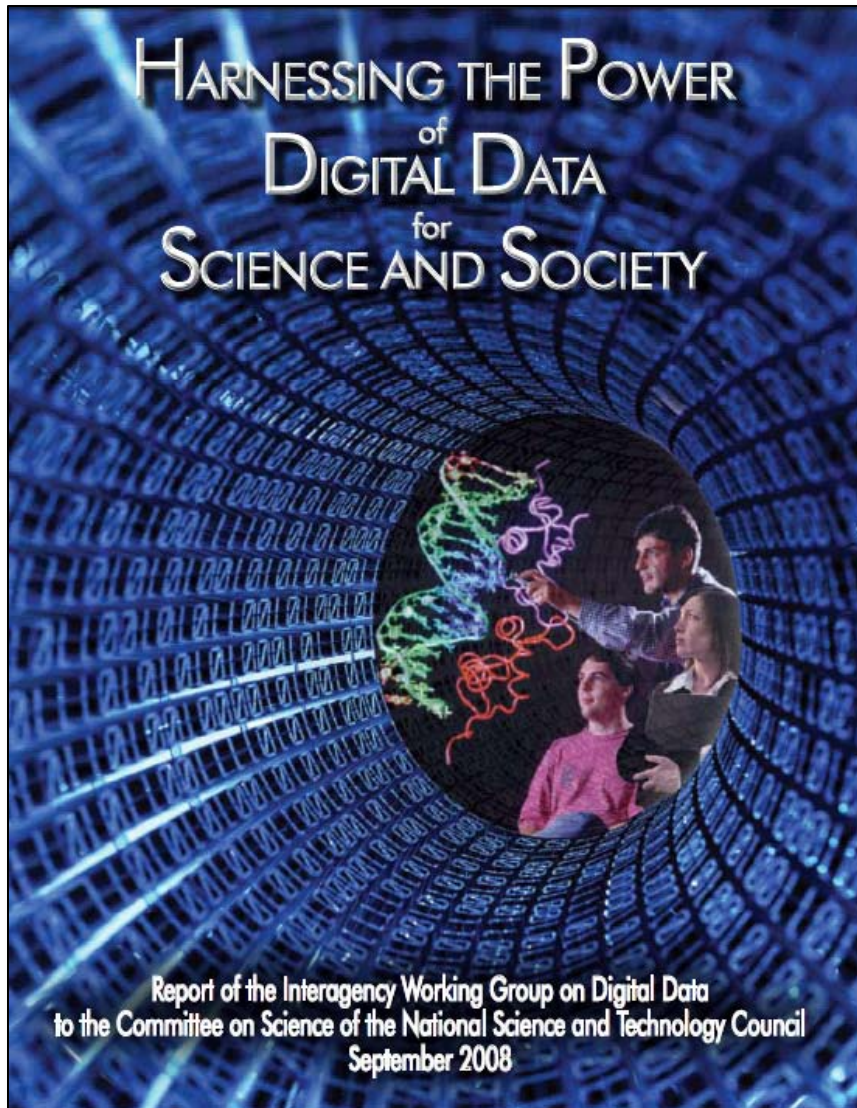
Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See [Grant Proposal Guide \(GPG\) Chapter II.C.2.j](#) for full policy implementation.

Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in [Grant Proposal Guide, Chapter II.C.2.j](#) apply.

Please note that if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed.

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



RECOMMENDATION 3 - DATA MANAGEMENT PLAN ELEMENTS

The third key element of the strategic framework is for all agencies to promote a data management planning process for projects that generate preservation data. This includes preparing a data management plan in proposals for activities that will generate digital scientific data. Examples of elements that should be considered in such a data management plan are listed below. This listing can be consulted by agencies in developing an appropriate portfolio of specialized data management policies, with each policy crafted for the community and context in which a particular project or projects will be conducted. Each specialized policy may include or omit any of the elements listed below or add others as appropriate to the particular application or context.

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. Where standards are missing or inadequate, alternate strategies for enabling data re-use and re-purposing should be described, and agencies should be alerted to needs for standards development or evolution.

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. These resources and capabilities should be appropriate for the projected usage, addressing any special requirements such as those associated with streaming video or audio, movement of massive data sets, etc.

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.



DISCOVER

FACILITATING DATA-DRIVEN SCIENCE AT CORNELL

[Home](#)

[About](#)

[Documents](#)

[Related Links](#)

[Services](#)

[Contact](#)

[News](#)

The DISCOVER Research Service Group (DRSG) is a partnership between domain scientists, the Center for Advanced Computing (CAC), the Cornell University Library, and Fedora Commons. Sponsored by the Office of the Vice Provost for Research, the DISCOVER team facilitates collaborations among research groups, fosters cross-disciplinary analysis of data using an array of data mining and visualization tools, and supports the development of a robust cyberinfrastructure for data-driven science at Cornell.

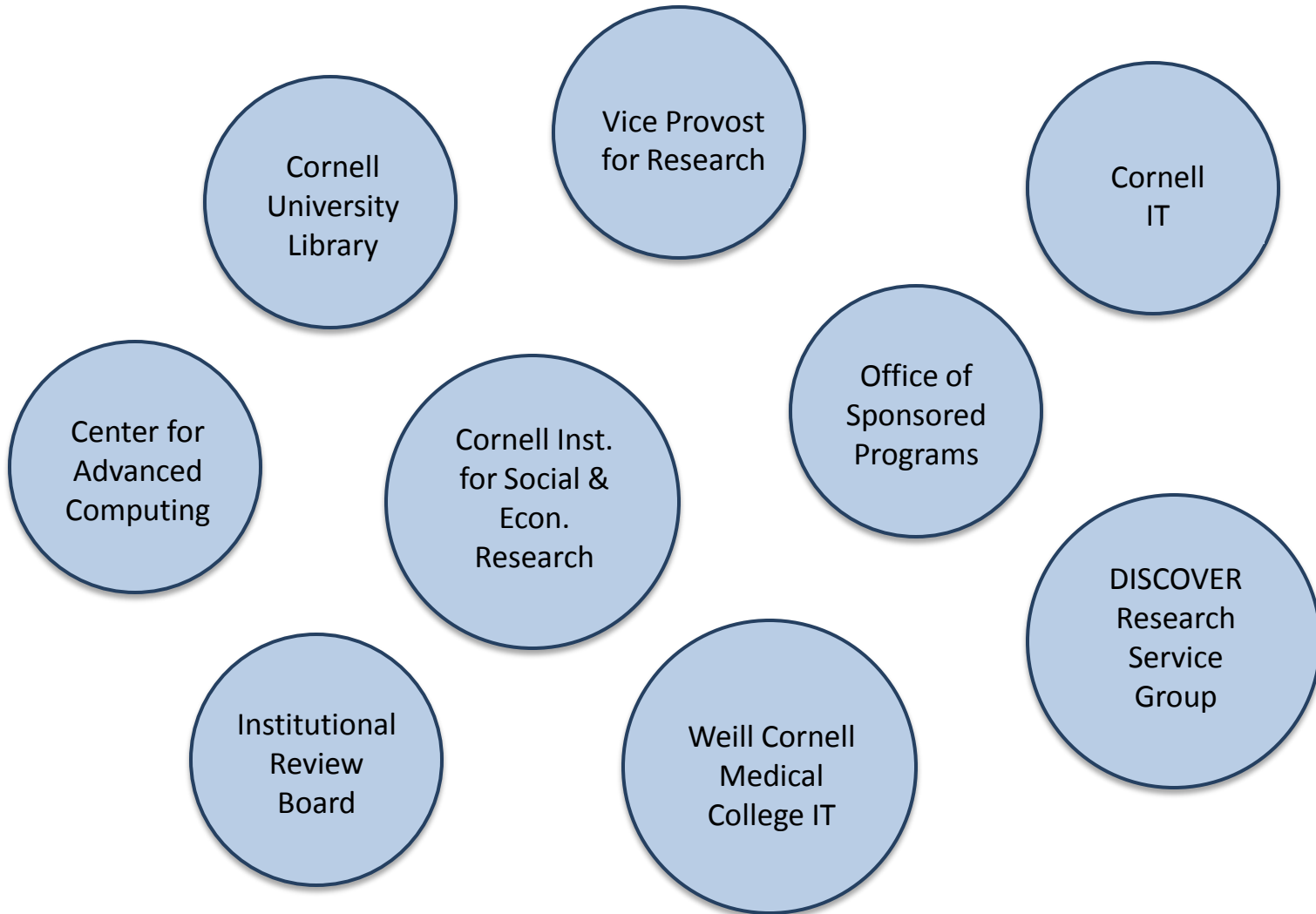
Robert A. Buhrman, Senior Vice Provost for Research at Cornell, says that "addressing the magnitude of data being generated by today's large-scale research programs is essential. That's why Cornell is expanding our investment in data-driven science by launching a new research service group that is dedicated to systematically tackling some of these data challenges." Because assimilating data into knowledge is typically more challenging than acquiring the data, the DISCOVER Research Service Group will work to provide accessible paths for the curation, preservation, and mining of scientific data. This more synergistic approach will avoid repetition, promote the adoption of best practices, and be more efficient and affordable.

DISCOVER is a partnership among domain scientists, digital library and data specialists, and colleagues in the Center for Advanced Computing (CAC). The purpose of the DRSG is to determine and respond to the needs of research groups in particular science areas; accordingly a number of domain groups are involved initially and we expect the number to grow.

Cornell Departments and Units

- [Astronomy](#)
- [Cornell Center for Advanced Computing](#)
- [Cornell University Library](#)
- [Crop and Soil Sciences](#)
- [Earth and Atmospheric Sciences](#)
- [Information Science](#)
- [Laboratory of Ornithology](#)
- [Physics](#)
- [Theoretical and Applied Mechanics](#)

<http://drsg.cac.cornell.edu/>





Research Data Management Service Group

[About](#)

[Resources](#)

[Services](#)

[Contact](#)

About RDMSG

Our goal is to make it as simple as possible for researchers to obtain the services they require. The Research Data Management Service Group (RDMSG) aims to:

- Present a coherent set of services to researchers;
- Develop a unified web presence providing general information on data management planning, services available on campus, and standard language that may be used in data management plans in grant proposals;
- Provide a single point of contact that puts researchers in touch with specialized assistance as the need arises.

The RDMSG is jointly sponsored by the Senior Vice Provost for Research and the University Librarian, and also has a [faculty advisory board](#).

- [Learn more about the RDMSG](#)

RDMSG News

[January information sessions on NSF-required data management plans](#)

Submitted by emc256 on Tue, 01/04/2011 - 15:22

Cornell University faculty and staff are invited to attend information sessions on the new National Science Foundation (NSF) policy requiring that all NSF proposals include a data management plan. This policy goes into effect January 18, 2011. Staff from the recently formed Research Data Management Service Group (RDMSG) will explain the new requirement and describe how researchers can obtain assistance from the RDMSG in order to create effective data management plans.

The information sessions will be held:

- Thursday, January 13, 1:30-2:30 pm – G01 Biotech
- Tuesday, January 18, 9:00-10:00 am – 102 Mann Library
- Thursday, January 20, 12:30-1:30 pm – 312 Hollister

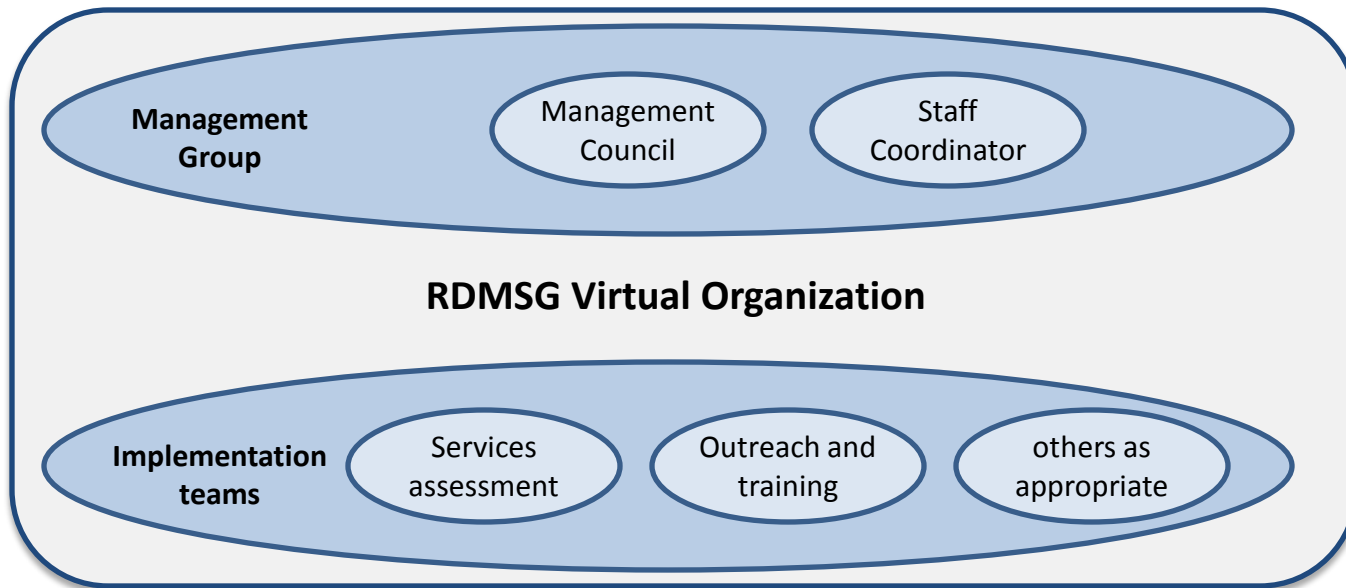


- Data management planning
- Storage and backup
- Metadata
- Data analysis
- Collaboration tools
- High performance computing
- Privacy and confidentiality
- Intellectual property
- Data publication

<http://www.flickr.com/photos/21709799@N03/3798486667/>

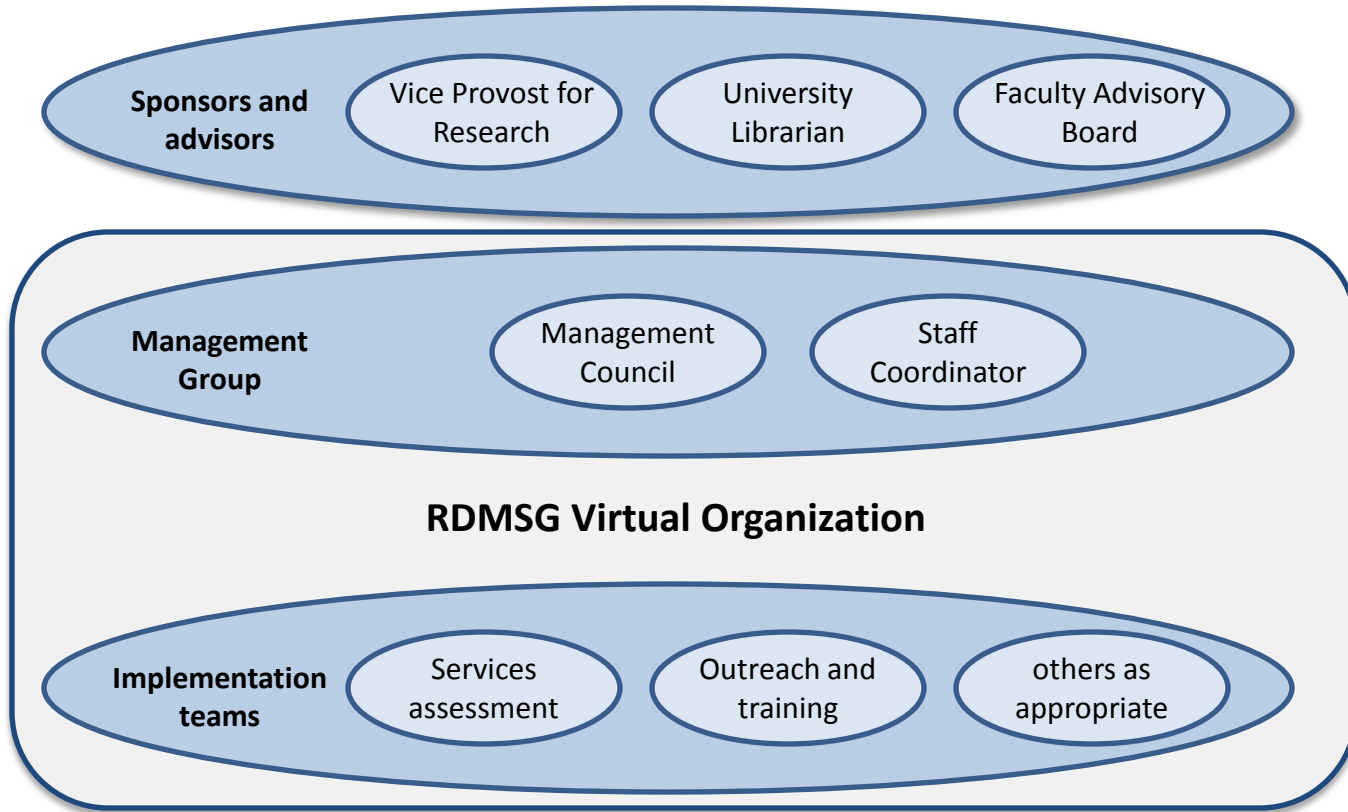


Research Data Management Service Group (RDMSG)



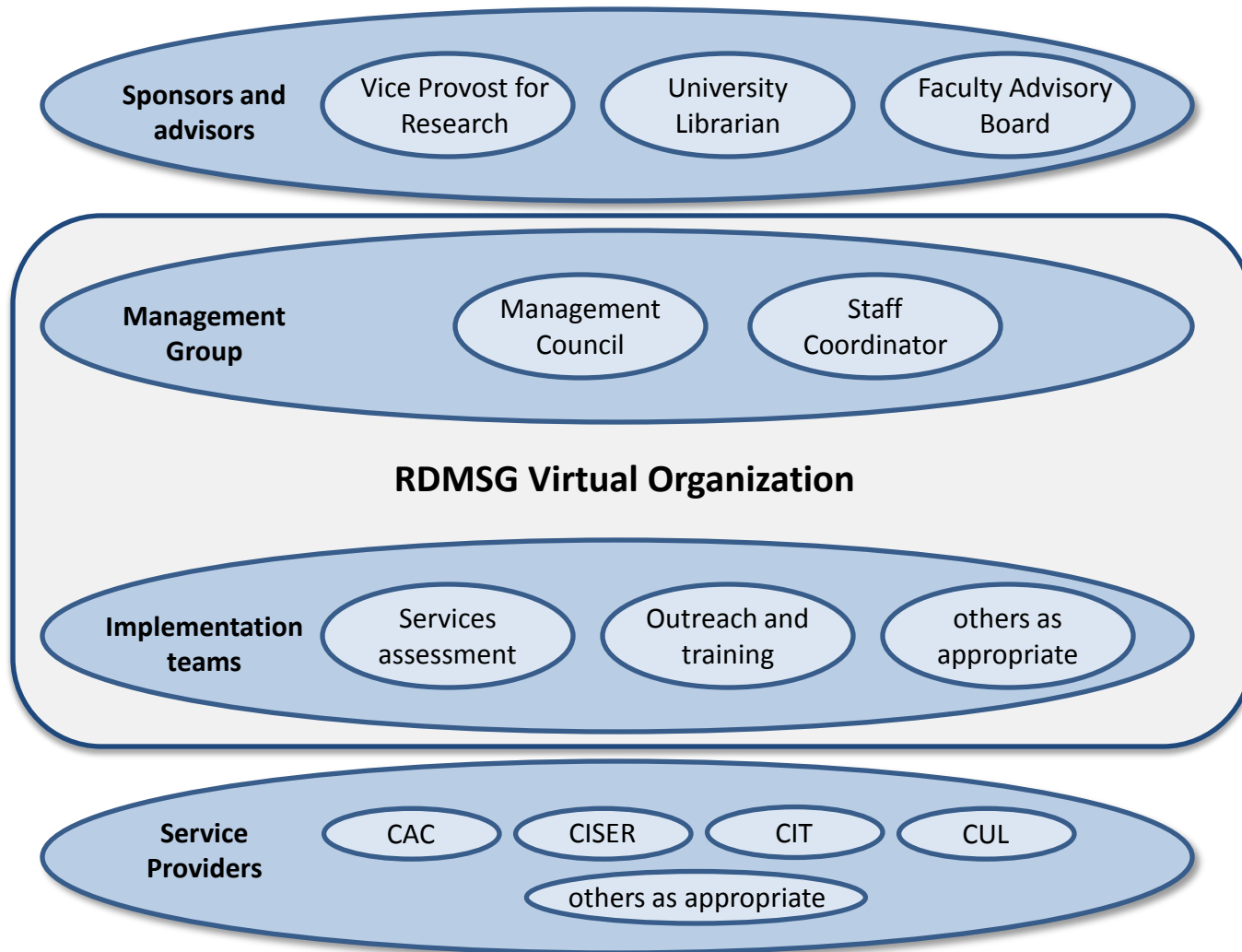


Research Data Management Service Group (RDMSG)





Research Data Management Service Group (RDMSG)





RDMSG activities

- Survey of active NSF PIs
- RDMSG website, rdmsg-help@cornell.edu
- Consultant pool
- Ticketing and triage system
- Information sessions on NSF DMPs
- Faculty advisory board
- Implementation teams:
 - Outreach and training
 - Documentation
 - Intellectual property and copyright
 - Survey report
 - Sharing and access
 - Long-term preservation



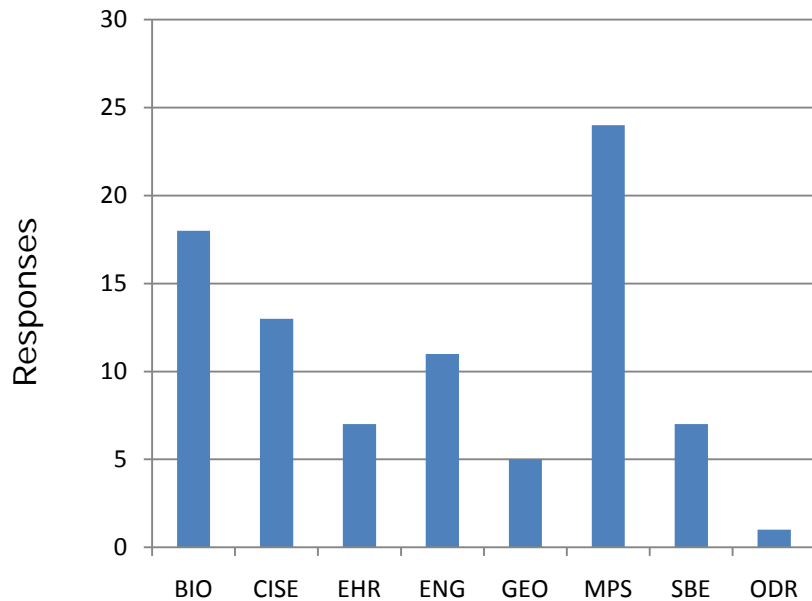
RDMSG activities

- Survey of active NSF PIs
- RDMSG website, rdmsg-help@cornell.edu
- Consultant pool
- Ticketing and triage system
- Information sessions on NSF DMPs
- Faculty advisory board
- Implementation teams:
 - Outreach and training
 - Documentation
 - Intellectual property and copyright
 - Survey report
 - Sharing and access
 - Long-term preservation

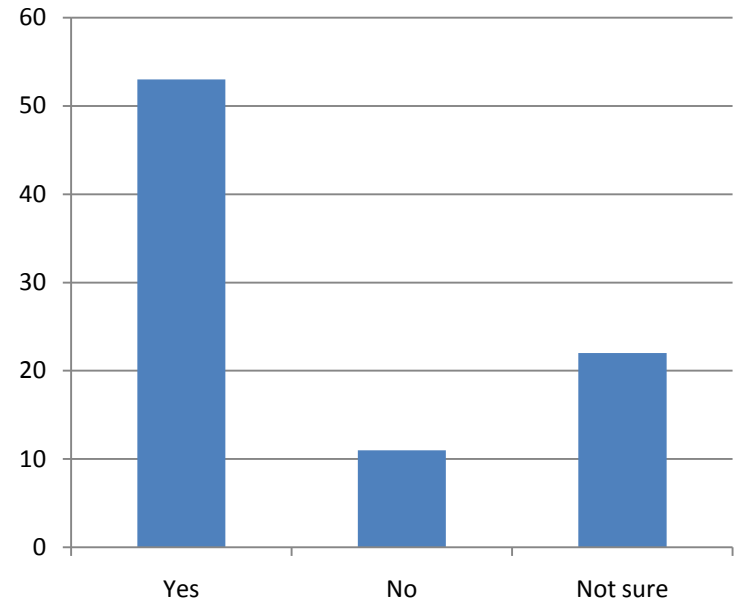


Survey results - general

Directorate of most recent award



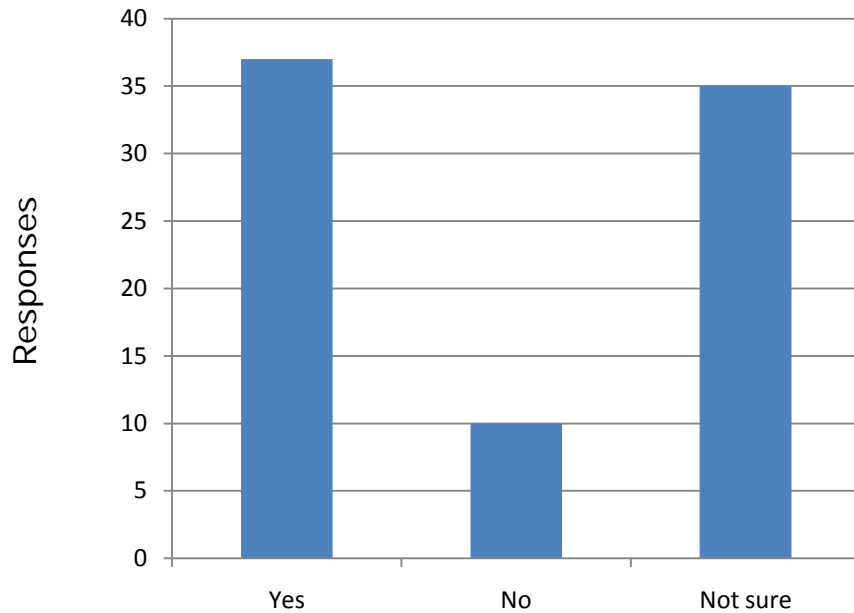
DMP support?



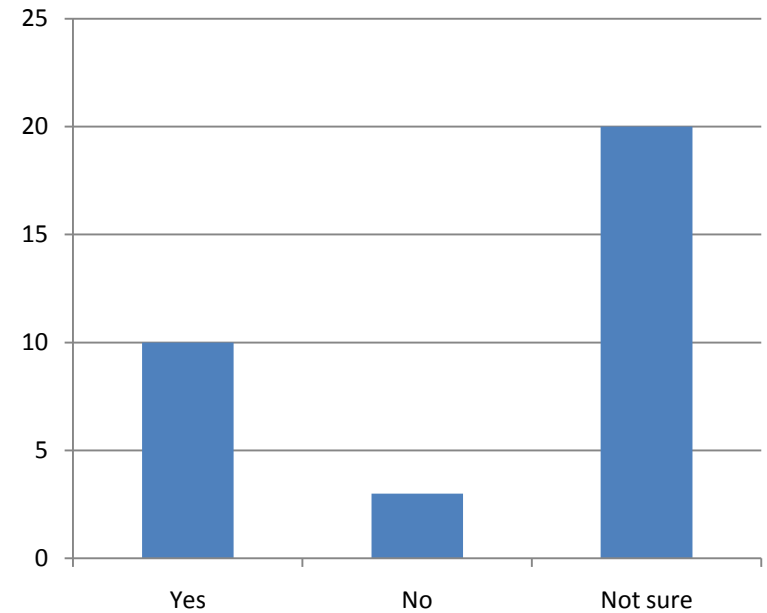


Survey results - standards

Data conform to disciplinary standards?

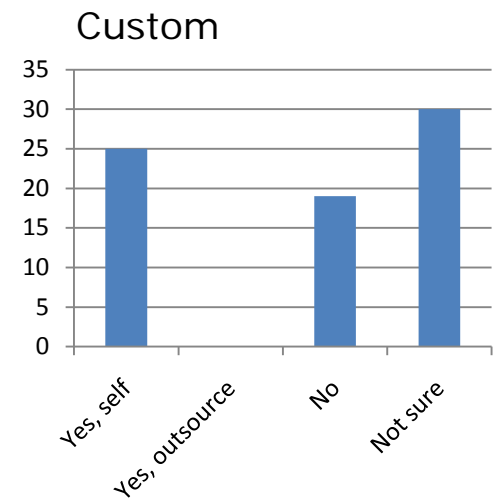
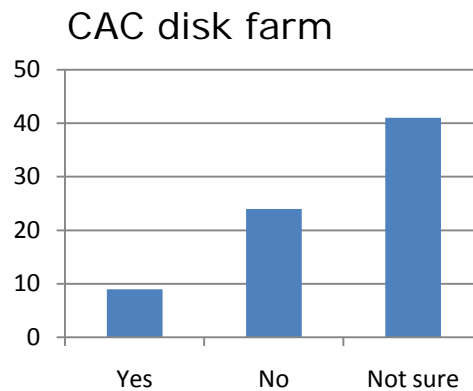
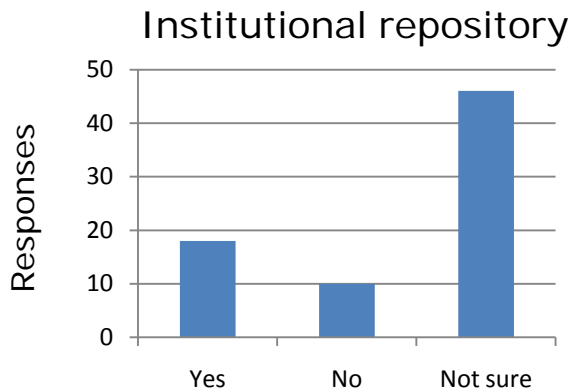
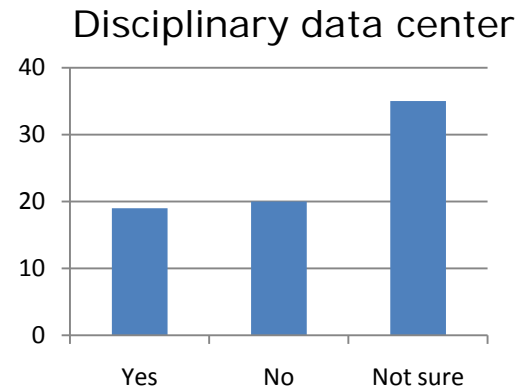
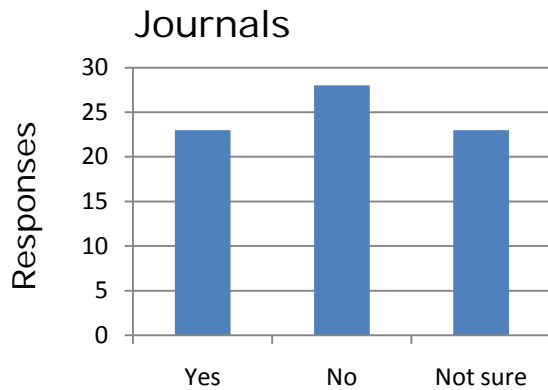


Metadata conform to disciplinary standards?





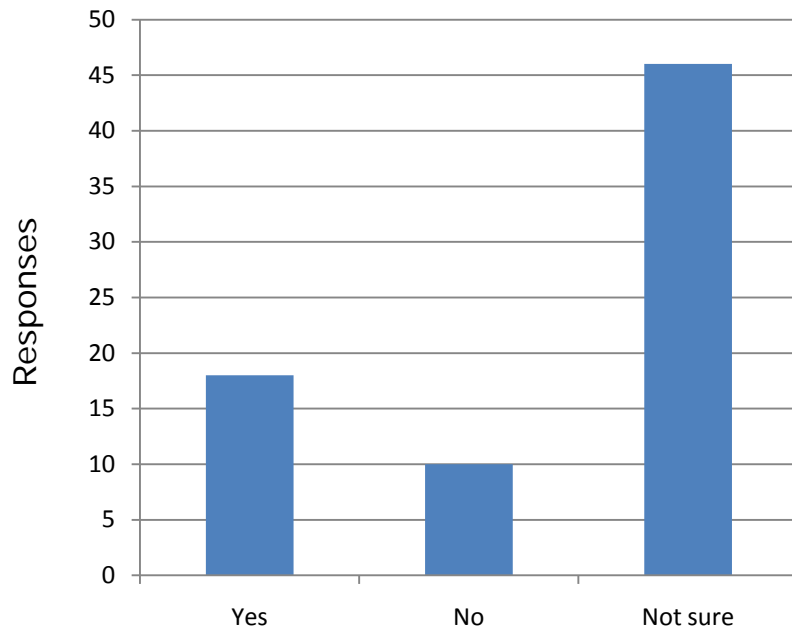
Survey results – sharing strategy



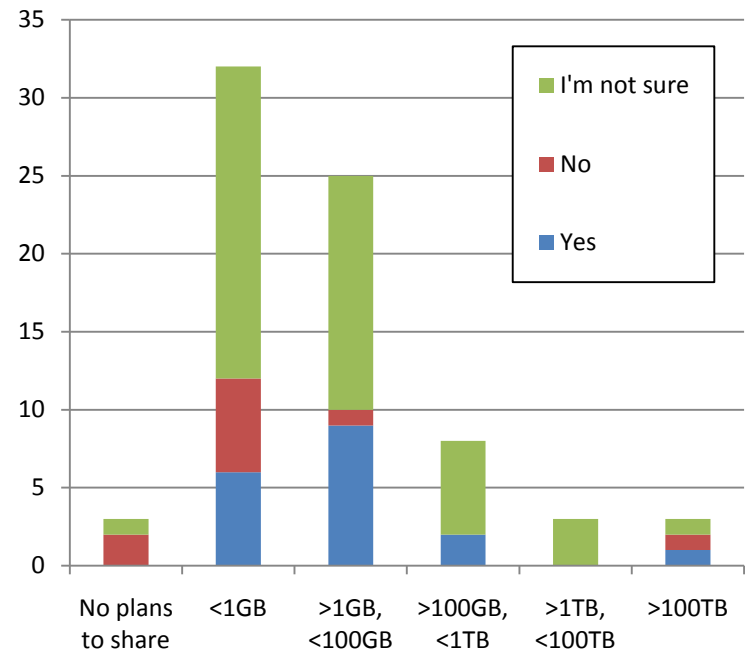


Survey results – sharing strategy

Institutional repository

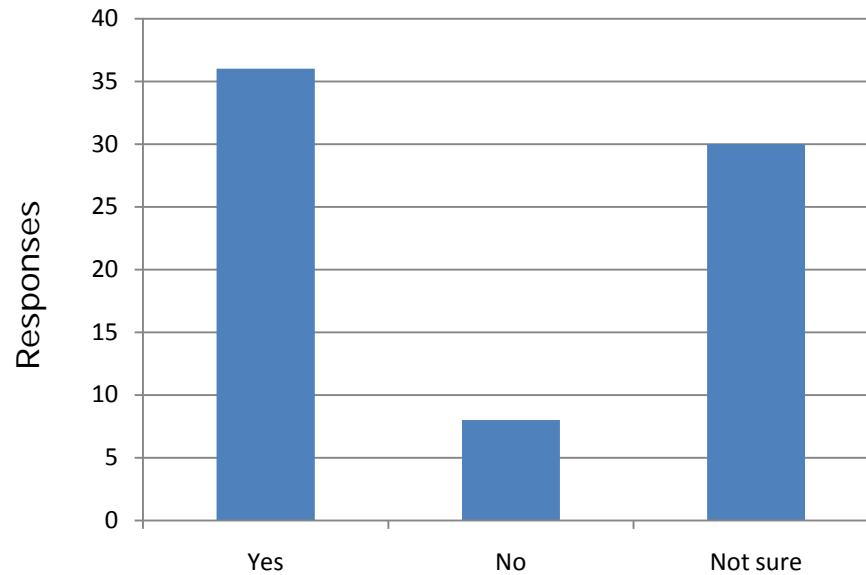


Intent to use IR by amount of data





Survey results – compliance/registry service





Info sessions: reactions

What counts as data? Does that include raw data?

I publish my findings in journals, so I already share my data.

How do we budget for data storage and access beyond the end of a grant?

For how long should data be archived?

What about the additional burden on reviewers?

What if your research plans change so that the original DMP is no longer relevant?



Challenges and questions... from us:

- Lack of detail makes support difficult
- Finite award period vs. longer term needs
- What services will researchers want/use?
- Potential for interim / less than optimal solutions



Gail Steinhart
Research Data & Environmental Sciences Librarian
Cornell University Library
GSS1@cornell.edu

Coalition for Academic Scientific Computation Meeting
March, 2011