



Cornell University's
Research Data Management Service Group
Summary of Activities and Accomplishments
July 2012 – June 2013

Contents

Introduction..... 1

RDMSG Outreach and Training Team 2

RDMSG Consultation Activities..... 5

RDMSG Documentation Team..... 7

RDMSG Researcher Survey.....10

RDMSG Publications and Presentations.....21

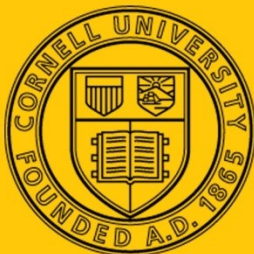
Summary and Conclusions22

Appendix A (Organizational Structure):.....23

Appendix B (RDMSG Faculty Advisory Board Members):.....24

Appendix C (RDMSG Management Council Members):.....24

Appendix D (Text of the RDMSG Researcher Survey):.....25



rdmsg-help@cornell.edu

data.research.cornell.edu

Introduction

After discussions in early 2008 addressing issues surrounding data discovery, re-use and curation, and a two year project funding by the Vice Provost's office to build campus-wide collaboration in support of data services, the Research Data Management Service Group (RDMSG) was founded in December of 2010.

The RDMSG is a cross-campus, collaborative organization that helps researchers find the data management services they need, and provides assistance in the preparation of data management plans for research funding proposals. Current service providers include: the Center for Advanced Computing (CAC), Cornell Information Technologies (IT@Cornell), the Cornell Institute for Social and Economic Research (CISER), Cornell University Libraries (CUL), and Weill Cornell Medical College (WCMC) (see also appendix A). The group receives oversight by a Faculty Advisory Board (Appendix B), and activities and services are developed and coordinated by the RDMSG Management Council (Appendix C).

This report contains a summary of activities of the RDMSG, from July 2012 through June 2013. Three implementation teams, Outreach and Training, Consultants, and Documentation, were active in this time, and this report includes summarizes their accomplishments, including the following highlights:

The outreach and training team focused efforts on specialized sessions, introducing faculty, staff and students to the RDMSG at 13 campus events, teaching 7 specialized data management-related courses or workshops, and holding 3 general DMP information sessions.

The RDMSG expanded their team of consultants to 14, broadening its range of expertise to include scholarly communication, digital asset preservation and health science data analytics.

Consultants held 27 individual consultations, assisting faculty and student with Data Management Plans for NSF proposals and other data management needs.

The RDMSG website (<http://data.research.cornell.edu>) had over 5,000 visits, with just over 1,700 of those being from Ithaca (33%) and over 3,300 (65%) being unduplicated (unique) visitors.

During the summer of 2013, the RDMSG distributed a survey to all Cornell affiliates who had submitted funding proposals since Jan of 2011. Respondents answered questions about their experiences with agency responses to Data Management Plans, satisfaction with their RDMSG interactions and provided suggestions for desired data services at Cornell.

RDMMSG Outreach and Training Team

Team members: Dianne Dietrich (Apr 2013 – present), Wendy Kozlowski (Dec 2011 – present), Gail Steinhart (lead; Dec 2010 - present), Sarah Wright (Apr 2011 – present)

Charge: The RDMMSG Outreach & Training team develops and delivers training materials for Cornell researchers on issues of data management, with particular reference to meeting the requirements of research funders, and to services available through RDMMSG providers. The Outreach & Training team also has primary responsibility for marketing RDMMSG services to the Cornell community. The Outreach & Training team strives to fulfill its charge by recommending best practices while also meeting the practical needs of Cornell's researchers.

Scope: The team's target audience is researchers at Cornell. The RDMMSG Outreach & Training team is not charged with training RDMMSG participants (consultants, service providers) or with handling communication tasks of the RDMMSG such as ongoing communication with (or recruitment of new) RDMMSG service providers.

Overview of Major Activities

- The Outreach team is working on new content for the RDMMSG website, including training materials on selected topics in data management, as well as a change in design. We anticipate completing this work by the end of 2013.
- Because the website has become a primary focus for the team, we've invited Dianne Dietrich (part of the web team; Sarah Wright is the other member and is already a member of the Outreach team) to join us on a regular basis.
- The Outreach team continued to hold information sessions, but due to a drop in attendance, we offered fewer of them. We also held sessions for more targeted audiences and participated in various outreach events around campus. See the list of outreach and training activities that follows.
- We began to experiment with walk-in consultations held at various locations around campus (more information below).

Summary of Outreach and Training Activities

- Center for Vertebrate Genomics Annual Symposium (7/12/12): Sarah Wright presented a poster. Attendance: not recorded.
- Show me the money (7/16/12): Participated in graduate student workshop on obtaining research funding. Olin Library. Attendance: 12.
- Managing Data to Facilitate your Research: Introduction to Data Management for Chemistry Graduate Students (8/2/12). Presentation by Sarah Wright. Attendance: 25.
- IT@Cornell – Planning Our Future Together (9/11/12): Presented a poster at Cornell IT community event. Attendance: ~300 for entire event.

- Inside Scoop (9/13/12): Participated in a library orientation event for graduate students. Attendance: ~90 for entire event.
- Introduction to RDMSG for Academic Technologies staff (9/25/12): Wendy Kozlowski presented an introduction to RDMSG and discussed possibilities for collaboration / referrals between the two groups. Attendance: 4.
- Grant Funding Databases (9/25/12): Elaine LaRocque was invited to teach a workshop at Mann Library and Gail Steinhart made a brief pitch for RDMSG at the beginning of the workshop. Attendance: 35 (approximate).
- NSF Data Management Planning Information Session (10/8/12). Attendance: 14.
- NSF Data Management Planning Information Session (10/9/12). Attendance: 8.
- Scholarly Communications for Library Liaisons – Lightning Session (10/10/12): RDMSG participated in an overview of scholarly communications services at Cornell. Attendance: ~15.
- Scholarly Communications for Library Liaisons – Lightning Session (10/15/12): RDMSG participated in an overview of scholarly communications services at Cornell. Attendance: ~15.
- Data-driven Scholarship for Library Liaisons (11/14/12): Introduced Cornell library liaisons to RDMSG. Attendance: ~15.
- New Faculty Data Management Planning Information Session (1/16/13): Information session for new faculty indicating interest in RDMSG on survey from the Office of the Vice Provost for Research. Attendance: 1.
- Managing Data to Facilitate Research (NTRES 6940, Spring 2013): Sarah Wright collaborated with Natural Resources Faculty member Cliff Krafft to teach a 1-credit, 6-week course in data management, geared towards students in Natural Resources and related fields. Enrollment: 25-30. See also “*Press*” below for a Cornell Chronicle article on the course.
- RDMSG round table with library staff (3/12/13): Provided library staff with a brief update on RDMSG and held an open question and answer session. Attendance: 8.
- Apparel Design Grad Student Seminar (5/1/13): Sarah Write presented an introduction to data management for graduate students. Attendance: 12.
- IT@Cornell (6/13/13): Presented a poster at Cornell IT community event. Attendance: ~250 for entire event.
- Excel Basics Workshop (6/25/13): Sarah Wright and Wendy Kozlowski (along with Gaby Castro-Gessner of CUL) taught a 1.5 hour workshop on best practices, tips and tricks for using Microsoft Excel to work with research data. Attendance: 7.

Walk-in Consultations: In an effort to keep RDMSG in the public eye and to make our services more visible, we began to offer walk-in consultations in March 2013 (see <http://data.research.cornell.edu/calendar/> for the schedule and locations). We advertised these via LCD displays located in Mann Library, Duffield atrium, Physical Sciences, and Space Sciences. Consultants are stationed in public, visible areas, and display an RDMSG sign while on duty. The participating consultants are Wendy Kozlowski, Sarah Wright, Dianne Dietrich, Adam Brazier, Gail Steinhart, and Florio Arguillas.

Approximately 8 hours of walk-in time are scheduled per month, at the convenience of participating consultants and dependent on the availability of consulting spaces.

Interest has been very limited, with only 3-4 walk-in consultations since we began, although Florio reports doing many CISER consultations during his scheduled walk-in hours. We plan to continue offering walk-in hours through the fall semester, and then to revisit whether to continue. Discontinuing them would mean taking down the LCD displays, and we speculate that this is good advertising even if uptake has been disappointing. We'd need to find other ways to publicize RDMSG.

Marketing and Communication: Our marketing and communication efforts included web-based communication (news items on the website, Twitter), LCD signs to promote walk-in hours, engaging the library's liaisons to recognize RDMSG referrals among the faculty with whom they work, a Cornell Chronicle article, and presentations and publications beyond Cornell.

Plans for 2013 - 2014

- Continue to develop content for the RDMSG website, as needed, including funder specific information, a reorganized set of guidelines for data management plan preparations and service provider spotlights.
- Continue to communicate with CUL liaisons to encourage referrals to RDMSG.
- Monitor emerging policies of funders such impacted by the Office of Science and Technology Policy memo on open access to research outputs, including publications and research data, and plan additional outreach accordingly (information sessions, Chronicle article, etc.).
- Engage in other outreach opportunities as they arise. Explore with OSP, ORIA and other campus groups opportunities for interaction with faculty, staff and research administrators.

RDMMSG Consultation Activities

Consultant Team Members: Florio Arguillas (May 2012 – present), Aaron Birkland (Dec 2011 – present), Daina Bouquin (Jul 2013 – present), Adam Brazier (Dec 2010 – present), Dianne Dietrich (Dec 2010 – present), Peter Hirtle (Mar 2011 – present), Keith Jenkins (Oct 2011 – present), Wendy Kozlowski (coordinator; Dec 2011 - present), Wyman Miles (Oct 2011 – present), Michelle Paolillo (Mar 2013 – present), Jaron Porciello (Mar 2013 – present) Gail Steinhart (Dec 2010 – present), Drew Wright (Dec 2011 – present), Sarah Wright (Apr 2011 – present)

Operating Principles:

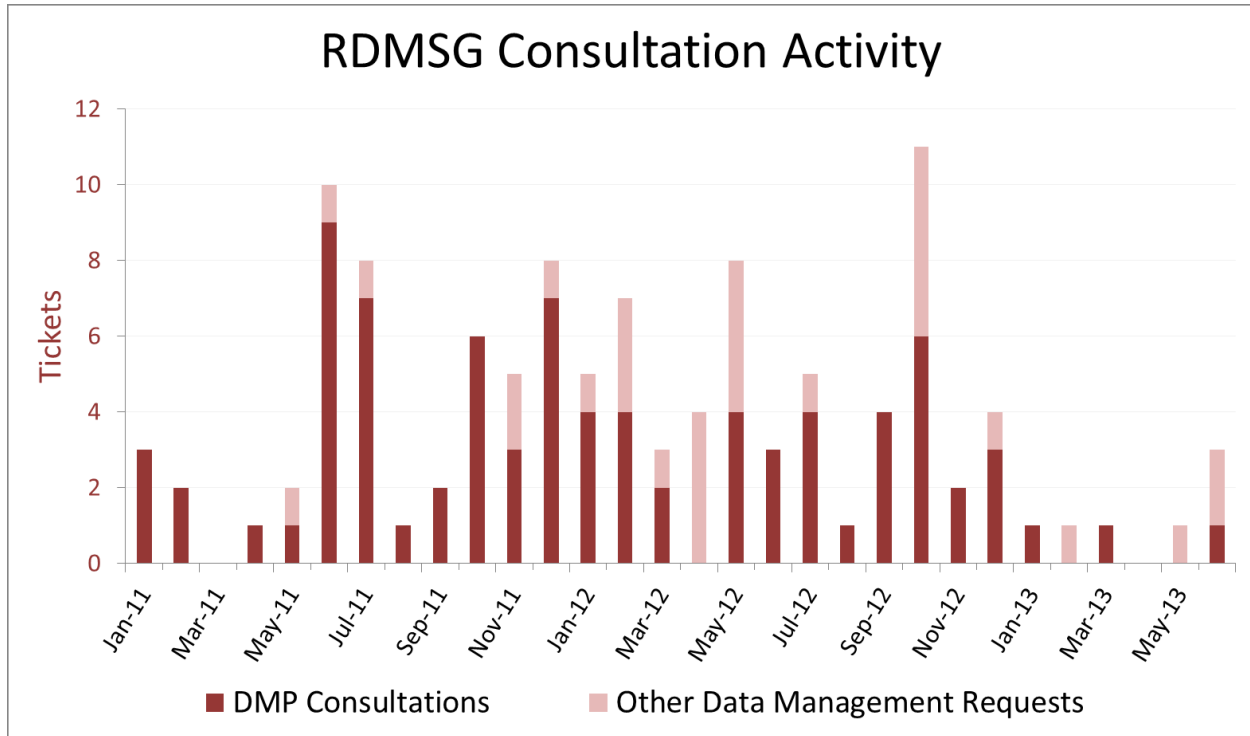
RDMMSG consultants strive to meet the Cornell research community's data management needs by:

- Providing timely and professional assistance in the creation and implementation of data management plans.
- Encouraging best practices in data management, including those that promote sharing, reuse, and preservation of research data, while respecting the concerns and practical constraints researchers face.
- Bringing diverse backgrounds and expertise to the RDMMSG and recognizing when additional expertise may be required to meet a researcher's needs.
- Sharing information with other RDMMSG consultants to provide the best possible service, while collectively treating all information in grant proposals as confidential.
- Referring researchers to the most appropriate services, whether at Cornell or elsewhere.

Summary of Activities, July 2012 – July 2013

- Met regularly to share experiences and information regarding consultation issues, as well as developments at Cornell and elsewhere related to data management.
- Three new consultants were added to the group; Jaron Porciello and Michelle Paolillo from CUL bring expertise on scholarly communication and repository management, and Daina Bouquin joins Drew Wright from the WCMC Library in supporting Weill researcher's data management needs.
- Developed a quick-reference document for consultants about use of the RT ticketing system.
- Customized a Data Management Plan-preparation tool (DMPTool: <http://dmp.cdlib.org>) produced by the California Digital Library System for use by Cornell researchers. Accessible via Shibboleth login using a Cornell NetID, this tool provides a working space for DMP planning and preparation with funder-specific guidance on content and links to relevant Cornell-specific resources.
 - The RDMMSG has not largely marketed this tool. There are some cases where use of the DMPTool is appropriate but we feel that the DMP preparation guidance provided on the RDMMSG website provides an easily-understood, extensible framework data management planning.
- 37 inquiries were received via the RDMMSG RT ticketing system (rdmsg-help@cornell.edu) between July 2012 and Jun 2013. These led to 27 individual consultations with Cornell faculty, staff and students.
 - Similar to the first 18 months of RDMMSG activity, nearly all consultations (~85%) involved questions directly related to NSF's Data Management Plan requirements.
 - Average time spent on consultations remains close to one hour (~55 minutes).

- Most consultations continue to be done via email (~76%), with the remainder being done face-to-face or by phone.
- Non-DMP-related consultations included conversations with institutions outside of Cornell regarding organization and logistics for data management support and outreach, electronic lab notebook support, metadata creation inquiries and copyright issues related to data collection.



Plans for 2013 - 2014

- Continue consultation efforts in support of agency data management plan preparation and other data-related needs.
- Remain abreast of possible changes in data (and publication) sharing requirements by funding agencies that may arise as a result of the Office of Science and Technology Policy memo of February 2013.
- Meet regularly as a group to share consultation experiences, and explore and educate ourselves about new tools and services at Cornell, Weill Medical or elsewhere.
- Consider and review Data Management Plans “donated” from Cornell researchers and other institutions as appropriate, to potentially develop a browseable library of DMP samples.

RDMSG Documentation Team

Team members: Dianne Dietrich (lead; Dec 2010 - present), Sarah Wright (Apr 2011 – present)

Charge: The RDMSG Documentation team has primary responsibility for maintaining the RDMSG website. The Documentation team will also work closely with RDMSG teams on external and internal documentation needs.

Scope: The RDMSG website will include information for researchers about funding requirements, service providers, the RDMSG team, and other information as appropriate. The target audience includes researchers at Cornell and RDMSG consultants. The Documentation team will work with any appropriate Cornell-related groups, including the RDMSG implementation teams, web design teams, all as appropriate.

RESEARCH
Data Management
SERVICE GROUP
comprehensive data management planning and services

Home Guidance Documents Services Contact About FAQ Search

Need to Write a Data Plan?

- [Writing a data management plan](#)
- [Frequently asked questions about NSF data plans](#)
- [Slides from our NSF data plan info sessions](#)
- [Data management guidance](#)
- [Consultant office hours](#)

Services for Managing Data

- [Storage and backup](#)
- [Intellectual property and copyright](#)
- [Data publication](#)
- [Metadata](#)
- [Data analysis](#)
- [Collaboration tools](#)
- [High performance computing](#)
- [Privacy and confidentiality](#)

Recent News

- [Mann library offering data management workshops](#)
Mann library is offering a series of workshops on data-related topics this fall. [\(Read more\)](#)
- [OSTP Information on Open Government Data](#)
The Office of Science and Technology Policy has released more information regarding "[Progress Toward Opening Government Data Resources](#)". [\(Read more\)](#)
- [Electronic Lab Notebooks - From Pilot Program to Site License](#)
[\(Read more\)](#)
- [All news](#)

Need help right now?

[Contact us](#),
check our [office hours](#),
or consult our guide, [Writing a data management plan](#)

Tweets

Follow

Cornell RDMSG 18 Sep
@CUrdmsg
IRB on-campus help session Thur, 9/26, 3-5pm, 303 Olin Library. Sessions are open, drop-in format. Full schedule: <http://www.irb.cornell.edu>

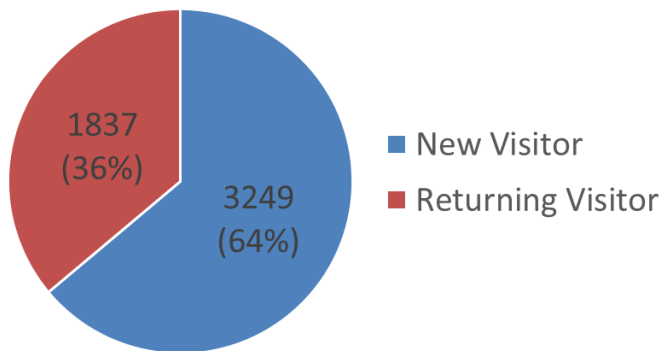
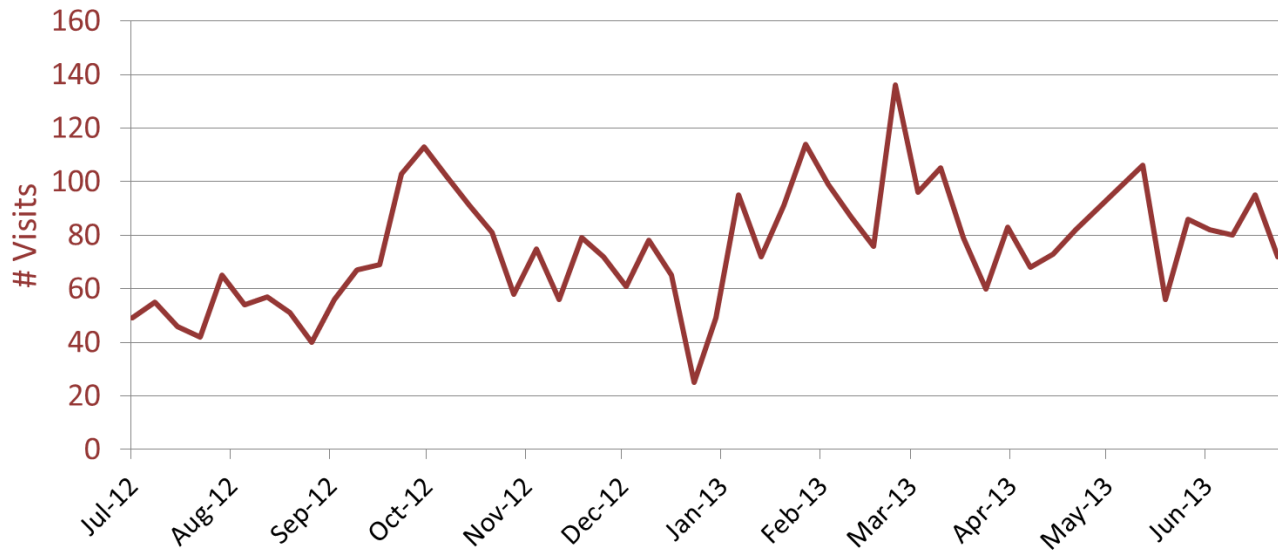
Cornell RDMSG 17 Sep
@CUrdmsg
Mann Library offers a series of data management workshops for grad students: confluence.cornell.edu/x/44L
Expand

Cornell RDMSG 17 Sep
@CUrdmsg
Cloud curious? CAC announces free exploratory Red Cloud accounts:

Documentation Team Activities

- Merged with Outreach & Training team in April 2013. The Documentation Team will continue to meet separately as needed for site maintenance and other technical needs.
- Developed a timeline for ongoing communication with RDMSG Coordinator, Management Council and Implementation Teams to ensure internal and external documentation needs continue to be met. (Note: This timeline has been tabled since the merge with Outreach and Training in April 2013.)
- **Maintained RDMSG website**
 - Continued to serve as contact for updates, edits, and any other changes to the site.
 - Developed strategy for the organization and presentation of material on the RDMSG website by defining audiences and reviewing counterpart pages from other institutions. (Note: After April 2013, this responsibility will belong to Outreach and Training.)
 - Created an automated weekly report that provided the status of all service provider links on the RDMSG website to more quickly fix broken or outdated links.
 - Coordinated review of Service Provider links; updated links when needed.
 - Organized and updated content, in consultation with RDMSG Implementation Teams.
 - In collaboration with the Cornell Libraries Usability team, did a cognitive walkthrough of the RDMSG website to get feedback on site organization.
 - Surveyed peer institutions' data management sites to inform our own content and presentation of site materials.
 - **Added and refined information for researchers:**
 - Added new content for funder-specific information
 - <https://confluence.cornell.edu/display/rdmsgweb/funder-specific-information>
 - Added yellow "Recommended Resources" boxes to longer pages to spotlight important site resources
 - Revised "Writing a Data Management Plan" to make it less focused on NSF
 - Created a "Guidance Documents" page to aggregate informational pages
 - <https://confluence.cornell.edu/display/rdmsgweb/how-to>
 - Added links to pertinent new tools on "Preparing Tabular Data for Description and Archiving"
 - <https://confluence.cornell.edu/display/rdmsgweb/preparing-tabular-data-description-and-archiving>

http://data.research.cornell.edu
Website Visits (weekly)



Plans for 2013 - 2014

- Implement proposed changes to RDMSG site, including new organization and layout.
- Continue to maintain RDMSG site by monitoring external links and implementing refinements to improve usability and discoverability of content.
- Outreach & Training will assume responsibility for content decisions and planning responsibilities for RDMSG website.

RDMSG Researcher Survey

Background: On June 6th, 2013, a survey was distributed to all members of the RDMSG-announce-l@cornell.edu mailing list (membership to which is upon request by interested faculty, staff and students) and any Cornell researcher that submitted a proposal to *any* funding agency since January of 2011, when the NSF requirement for inclusion of proposal Data Management Plans went into effect. The Cornell University web survey tool (Qualtrics) was used for dissemination and collation of results, the protocol was granted exemption from IRB review by the Office of Research Assurance and Integrity, and the total distribution list included 1619 individuals. A text copy of the full survey is included in Appendix D.

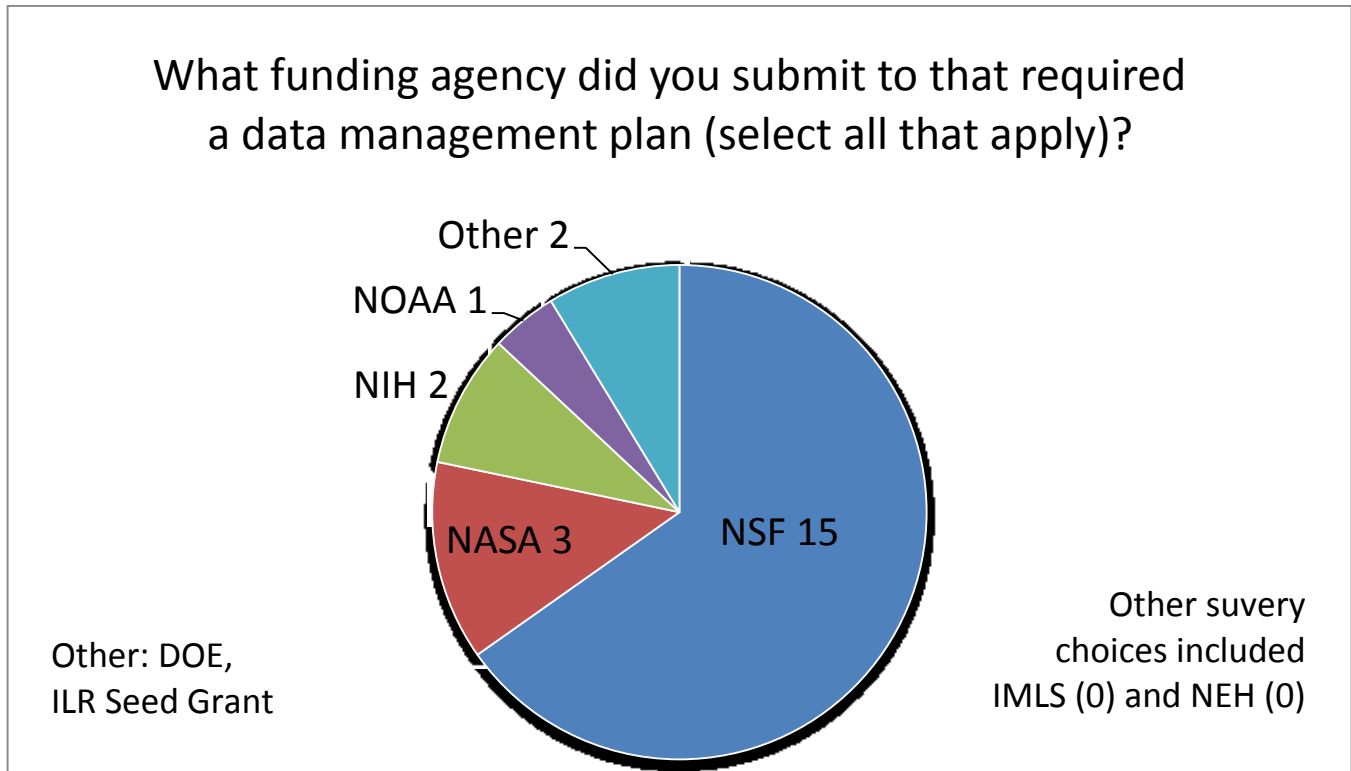
The goal of the survey was to gather information in three main areas:

1. Researcher experience with review of proposal Data Management Plans, including feedback received on submitted plans as well as experiences from serving on proposal review panels.
2. Resources cited to be used as part of a data management plan, including but not limited to Cornell-provided services.
3. Experience and satisfaction with the RDMSG and suggested areas for addition data management.

156 respondents completed the survey, translating to a response rate of 9.6%. An additional 57 respondents started the survey but never completed it. Summarized below are responses to all eight questions.

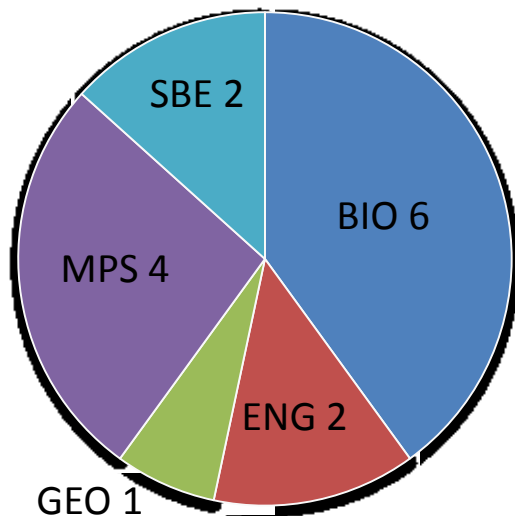
Question One:

What funding agency did you submit to that required a data management plan (select all that apply)?



Question Two:

IF NSF was selected in previous question - which NSF office or directorate was applied to (select all that apply)?



Other suvery choices included ERE (0), CISE (0), EHR (0) and Office of the Director (0)

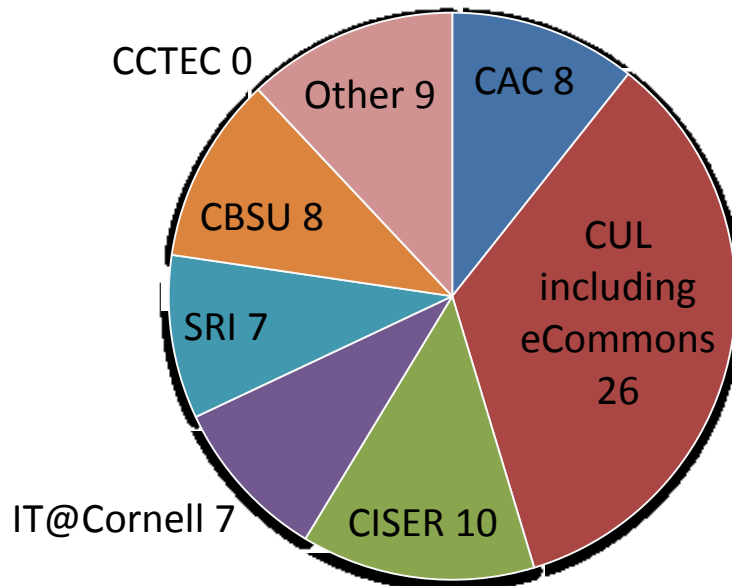
Question Three: If you received feedback on the DMP portion of your proposal, please share it here.

Simple or minimal feedback (11)	Have DMP experience (3)	Other (4)
I was informed that my DMP was "adequate".	We have been managing data on NASA work before the current requirements. Normally NASA HQ stipulates exactly what is required to collect as data. Sometimes this is tedious work. Like what are the students you supported on your grant doing now after you supported them for four years?	As I mentioned earlier, this is not so common for the proposals submitted to the NSF in the field of Harmonic analysis / and PDE (part of the DMS).
Very vague feedback such as "Data Management plan looks reasonable considering the size of the program."	Working in Planetary Sciences, we have been writing DMPs for, and deep-archiving data to, the Planetary Data System (http://pds.jpl.nasa.gov/) for many years, and there is a lot of experience in this area. In fact, since retiring from Cornell I have become a consultant in similar efforts, and am available for hire. / / Brian Carcich / 607-342-7595	This was for an international project, and our international collaborators from Zimbabwe were reluctant to make the data publicly available in the timeframe suggested by the NIH.
got no feedback	My data management plan followed the criteria used by the NSF Long Term Ecological Research (LTER) network of sites. Their plans are the result of considerable thought and experience. Hence my data management plans are accepted and approved.	I'm a historian [sic] of science. As such I use archives (unique sets of papers at rare book rooms in libraries) and take notes on them and sometime make xeroxes. So I am not generating any new data--the originals are still there and in fact one has to sign various documents before being allowed to use them. (One will cite them, one will seek permission to publish them, if one does, etc.) Sometimes I generate a table of data from published sources (lists of names, for example). Someday all my notes will go to the Cornell

Simple or minimal feedback (11)	Have DMP experience (3)	Other (4)
		University Archives. Therefore I did not think that I needed to do anything about NSF's new data management rules. They just do not fit my situation. But the NSF official disagreed, and I was declined. By now the success rate is so low (20% or less) and there was no encouragement to reapply, so I haven't.
our plan was deemed sufficient		CHESS award got feedback from DMR. That information should have been submitted to you by Joel Brock.
no weaknesses noted--received a perfect score		
Very brief comment that "Data Management Plan is satisfactory" or words to that effect.		
Our plan was rated as adequate		
I've received no feedback		
We were missing some information. Once we clarified, all was OK. There was some confusion and minimizing the topics covered has helped.		
My data management plan is pasted below. They thought the plan was appropriate and I was funded. / / "No special data management plan is needed for this work. Data will be disseminated through the normal mechanisms of seminars and publications. Strains will be stored in our laboratory in a -80C freezer and sent out as requested. Any E. coli strains that are produced that are of broad interest will be submitted to the NSF funded E. coli genetic stock center at Yale University. http://cgsc.biology.yale.edu/ "		
I'm actually unclear here whether DMP refers to a new, specific section or is more generic, referring to the kind of section on data management that one has included in proposals for decades. It's the latter that I'm responding about. / / With that context, reviews typically make some mention of data management, typically saying it looks okay, occasionally asking for a clarification about protecting subject privacy and the like.		

Question Four:

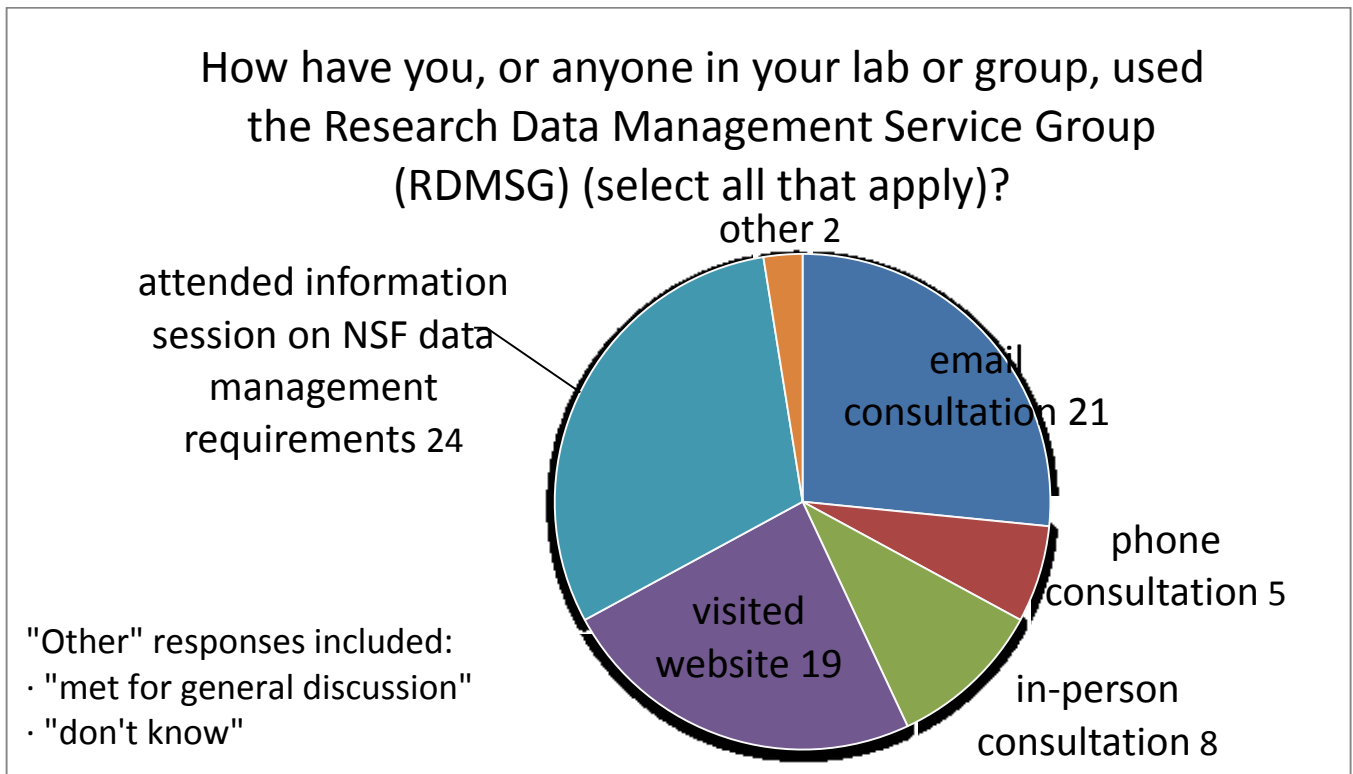
Which of the following Cornell Resources were cited in submitted DMPs (select all that apply)?



Free text responses to Question 4 included:

- NASA Planetary Data System
- NSI grants that I submitted were collaborative, with lead PI at other institutions. So data management plan was submitted by lead PI.
- E. coli genetic stock center at Yale
- Center For International Studies
- Landscapes and Objects Laboratory
- Shoals Marine Laboratory
- Bailey Hortorium Herbarium at Cornell University and http://cupac.bh.cornell.edu/cgi-bin/cupac/CUPAC_nav.pl Part of the Plant Ontology project
- CNF, CCMR, Inst. of Comp. Sci. & Engr. (new)
- Our own group resources within the Center for Radiophysics and Space Research
- CSU

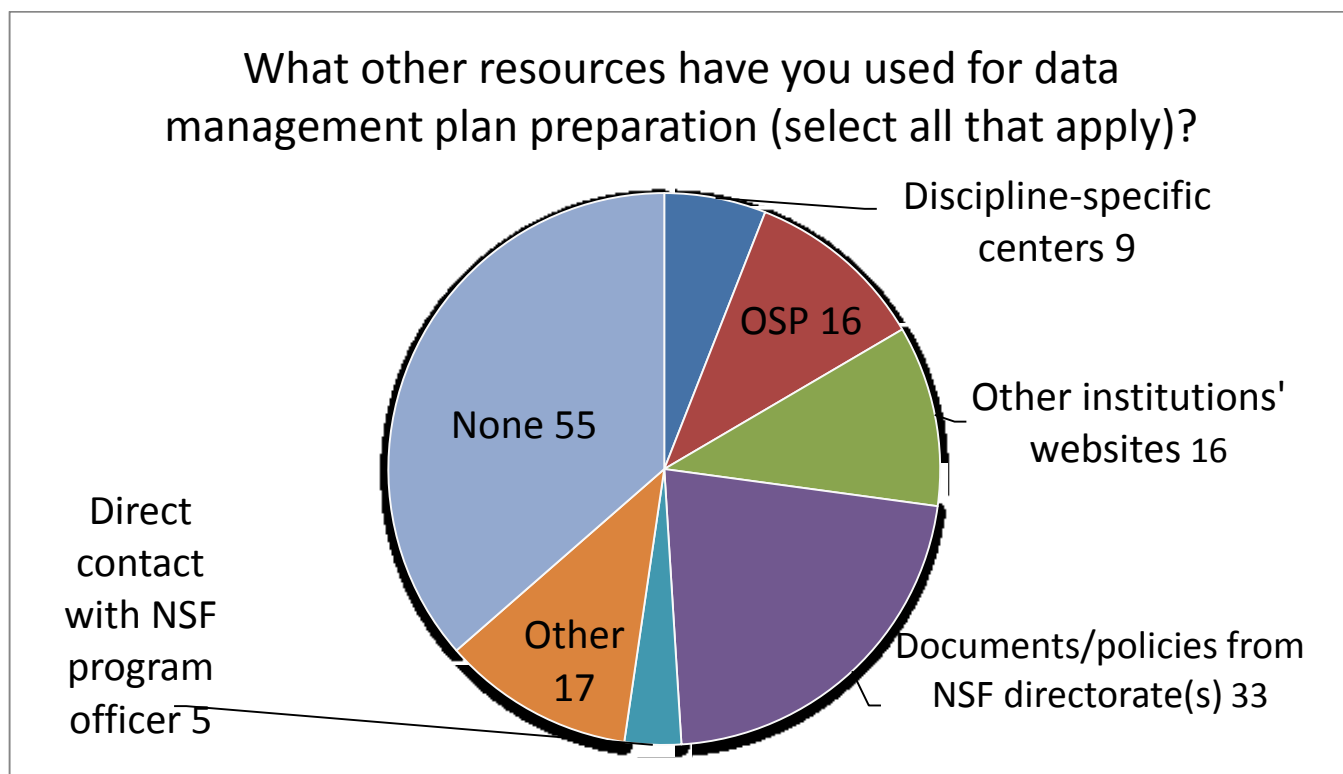
Question Five:



Question Six: How satisfied were you with your RDMSG interaction?

	<i>Very Satisfied</i>	<i>Satisfied</i>	<i>Neither Satisfied or Dissatisfied</i>	<i>Dissatisfied</i>	<i>Very Dissatisfied</i>
<i>Knowledge of relevant material</i>	19	15	3	2	0
<i>Professionalism</i>	21	16	2	0	0
<i>Promptness of response</i>	20	12	6	0	0

Question Seven:



Question Eight: Please use the space below for any suggestions of data management services that Cornell might expand or improve upon. [Please note that responses are grouped roughly by subject.]

Provide better storage/archiving solutions

- The single most important thing that Cornell could do would be to provide a service that provides indefinite storage for data after the lifetime of the grant. In the context of the data management plan, the NSF and reviewers are willing to accept significant costs to provide this service.
- It would seem to me that the University should develop a storage/access site for the collection of research data generated on Campus. This would then allow all Proposals to uniformly cite the CU resources available to meet these data management requirements.
- We are implementing our own solution for long-term mass storage. Our current system capacity is about 500 TB and we plan to expand it in the future. We find that we can do this more economically ourselves than through Cornell services. We would welcome a Cornell solution, but Cornell would need to provide a level of flexibility, eg for large volume and high speed, that is not currently available.
- We need data storage and web hosting to make image databases public. I'm going to try to do this on iplant, but we'll see if that works.
- Providing archiving services for data, software, and other materials.
- The data generated from research projects at Cornell is a material good that is produced at Cornell, and they should have a vested interest in preserving it via safe and accessible storage. This service

should be part of the regular Facilities and Administration Indirect cost rate that the university charges investigators. Losing data is a significant impediment to research advances. This is the key criticism (data storage/retrieval) that I have heard about in data management plans.

Increased outreach / communication about services available

- Tell us that these resources exist (which, in a backwards way, you just did).
- Could not find your group during Cornell related searches to meet data management requirements in proposal.
- New staff orientation programs across CALS, and periodic workshops.
- May be not Cornell provide data management service but provide help in identifying ways of managing data and what resources are available, etc.
- NIH also has requirements about a data sharing plan--is it possible that we can get that addressed in tandem with these NSF requirements, or at least consider that other federal agencies have requirements that are either current or on the horizon
- Most research is NIH-focused, unsure if there are people focusing on data management for NIH-type proposals.
- Cornell should make its faculty more aware of its data management services through workshops, because many faculty don't understand that data management means more than just saving files to a server.
- Provide staff to help faculty.

Data/Metadata/Tools Ideas

- I found the support helpful in writing the proposal - although it was still pretty abstract to me. As we kick off our project in a few weeks we will be needing to go back to the group for more concrete help. In particular understanding what metadata is, how to use it, how to generate it and how to follow the metadata standards is something that's basically a mystery to me. I think that it will be important for our project, but we still need to figure it out.
- It would be nice if data deposits communicated with each other, so that data could be deposited once but available from multiple portals.
- The university needs to provide actual services (not just cheap talk). There is no point in discussing how to archive and preserve data when none of the university resources can actually do that in an efficient way for (large and/or confidential) data. While the RDMSG has some useful links, it is mostly not useful for coming up with a real-life implementation. We need a resource at Cornell, or a collaboration with folks elsewhere at the University level, that actually provides (i) an actual long-term archive with full curation (ii) a tool/website/software that is easy to use (not custom programming each and every time) in the vein of the (Harvard) DataVerse. Even preserving *documents* is an exercise in frustration: neither eCommons nor DigitalCommons fully satisfy the minimal requirements I expect from a fully fledged document archiving site (à la JSTOR: actual ways to robustly cite resources; consistent and modern interface, etc.).

Review Panel Feedback

- I just served on an NSF panel review, in astrophysics, and I can promise that while DMPs may have started as a formality, that is changing quickly. The panel assessed the DMP of all proposals based on (1) backup plans, so that no data were lost, (2) archiving plans, so that data were preserved, and (3) plans to make the data available to the scientific community. How (3) should have been implemented

depended on the nature of the proposal, but generally, proposers were expected to either include the dataset as an appendix to a paper (which would then be available to the community via the journal and, most likely, VIZIER), or by posting it to an archive server directly, preferably one which would be around for the long term. Plans to host datasets on local web servers were generally not viewed favorably, due to the dependency of that approach on future funding. If Cornell is offering webhosting to PIs indefinitely, that should be made clear to all Cornell PIs.

Want DMP templates/examples or other DMP support

- Provide to administrative assistants who submit proposals in individual units appropriate templates for the specific agencies that they commonly deal with. For example, if a unit deals with NSF, Department of Defense and Department of Energy Proposals, then a template should be supplied for each of those organizations.
- I think providing templates for various types of data would be the most useful. Maybe you do that already, I didn't check.
- Provide a template. We use one from another university currently. We are not aware of a one stop shop and assume if there was one the info would not be kept current like the OSP IDC and f&A rates.
- CISER and Warren Brown alerted me to standard boilerplate language used to describe the data management system. It might be helpful if that were more widely circulated, as this process is time consuming and often confusing.
- It would be great to have standard procedures that we could plug into our proposals.
- Examples of successful plans, critiques of less successful plans
- Making available examples of successful data management plans for a range of disciplines and issues. / / 2. Providing "boilerplate" text that PIs could use or adapt to say "we will use [service X] at Cornell for [function Y] in our data management plan, and here's why that's a good idea that meets all our needs and all of NSF's requirements".
- It would be nice to collect plans and review comments from other colleagues so that we can see the feedback. My feedback during the review process was that my data management plans were adequate.
- The presentation on Data Mgmt plans was only marginally helpful. It would be helpful now that lots of Cornell DMPs have been submitted to find ones that someone in the OVPR thinks represent good examples for different areas (e.g. a couple of good ones for Engr.)
- There could be a template on the Cornell website. The person I worked with was incredibly helpful and helped me get it done with a really tight deadline. I appreciated that tremendously.
- Consider a web-based input form that elicits specific info (e.g. title of the project, how long the data should be kept, who should have access, etc) and then have a paragraph emitted which describes Cornell's promise to archive and serve the data with details provided from the answers to the question. The output just needs to be a short plain text paragraph. Most of us will just copy that into the proposal as needed, elaborating or trimming as deemed necessary.

Have not used RDMSG services / Not relevant to my work

- This survey seems to be irrelevant for my situation because I have no grants requiring data management other than what I normally do in carrying out research.
- did not have to submit data management plan ever so am not familiar with any of your questions

- Not relevant to my research. Just use the formula: / / The work proposed for support in this grant is entirely in pure mathematics. There will be no primary data, samples, physical collections used and no software or inventions developed. All research papers produced will be published in the usual way and/or distributed electronically by posting on the PI's website or others. Thus no other data management plan is necessary.
- I work in pure mathematics. We generally have no data to manage, and just submit a one-line planning saying so.
- to my knowledge, I have not submitted proposals that require data management plans.
- I have not used Cornell data management services and do not plan to use them. NSF and NASA already fund several data centers in my research area and these are the standards that we use
- I'm in the last month of an NSF grant, and don't intend to submit any other in the immediate future.
- My grant did not require a data management plan, as far as I know
- My data management plans have been very simple and straightforward. That is why I have not employed the resources offered by Cornell.
- I was not aware of many of the services offered /
- I have not had to use this service as we have not yet had a pre-proposal reach that stage.

Praise

- RDMSG was an extremely valuable and welcome resource.
- Note: did not submit proposal with DMP, so first question on this survey should include a "NA" response less the answers are misleading. Thank you for providing data management services, it is a great asset to have on campus.
- I found the support helpful in writing the proposal - although it was still pretty abstract to me. As we kick off our project in a few weeks we will be needing to go back to the group for more concrete help. In particular understanding what metadata is, how to use it, how to generate it and how to follow the metadata standards is something that's basically a mystery to me. I think that it will be important for our project, but we still need to figure it out.
- Nice short survey. Please request a survey again in ~6 months when I will have feedback on the specific data management plan that I proposed. Data management is yet another unfunded NSF mandate. The NSF requires us to make our primary data publicly accessible for years after the grant expires. Presently, there is no obvious way to bill the original grant for the archiving of the data. It would be helpful if Cornell set up a PAID service that would allow me to pay, up front, for 5 years of data archiving. This way I could charge the NSF for the data archiving costs before the grant expires. The NSF should bear the data archiving cost, not Cornell or me, through unrestricted funds.
- The help that we received for archiving research data was simply superb. As a result our data is now on the web and freely accessible to anyone.
- Things work very well now; I see no need for improvement (and I wish other parts of Cornell -- HR, accounting, OSP -- worked half as well).

Didn't know about you

- You need a better explanation of what this service is and how it works.
- I was unaware of any data management services at Cornell until I received this email. We proposed (to the NIH) a high-throughput sequencing project that would generate large amounts of data. I indicated that we would host this data on my laboratory website. This seemed to be satisfactory to the

reviewers. The grant scored well, but because of the current NIH budget funds still haven't been released.

- I don't remember my interaction with RDMSG very clearly.

Other comments

- keep it simple
- Thus far, it has worked well for me.
- The basic flaw of this survey is that you didn't start by asking if I had submitted a proposal since 2011 or whether it was an NSF proposal. Since I hadn't submitted a proposal, all of the questions were irrelevant.
- Use me as a resource. / / Develop standards for archiving various forms of data (again, <http://pds.jpl.nasa.gov/> is a decent starting point). I went to an early RDMSG information meeting, and most of the discussion was "What do I write in section X and Y of my DMP document to complete an acceptable grant proposal?" when a different focus set such as "What is a DMP?" or "How do we archive data?" would have answered the first question.
- Ought to get some clarification that taking notes on written materials already in an archive somewhere is hardly new data.
- Item # one does not allow for a skipped response. It was not applicable to me.
- In this survey, the first question should have had a "not applicable" for those of us who have not submitted grants that require data management plans or services.

RDMSG Publications and Presentations

- Kovari, Jason, Wendy Kozlowski and Danielle Mericle. 2012. Collaborative Service Models: Building Support for Digital Scholarship. Digital Library Federation Forum. November 5, 2012. Denver, CO. <http://hdl.handle.net/1813/34519>
- Kozlowski, Wendy. 2013. The Digital Research Lifecycle: Support and Engagement at Cornell University. American Library Association's Public Service Directors of Large Research Libraries Discussion Group. June 30, 2013. Chicago, IL.
- Kozlowski, Wendy, Sally Rumsey and Karen Smith-Yoshimura. 2013. Managing Research Data: From Goals to Reality. OCLC Research Library Partnership Webinar. May 14, 2013. <http://www.slideshare.net/oclc/managing-research-datafrom-goals-to-reality>
- Steinhart, Gail. 2013. How can we help you? Institutional Support for Data Management Planning. Columbia University Research Data Symposium. February 27, 2013. New York, NY.
- Steinhart, Gail; Chen, Eric; Arguillas, Florio; Dietrich, Dianne; and Kramer, Stefan. (2012). "Prepared to Plan? A Snapshot of Researcher Readiness to Address Data Management Planning Requirements." *Journal of eScience Librarianship* 1(2): Article 1. <http://dx.doi.org/10.7191/jeslib.2012.1008>

Summary and Conclusions

In February of 2013, the White House's Office of Science Technology and Policy director John Holdren released a memo directing "Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research."¹ The full text of the memorandum is available at http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

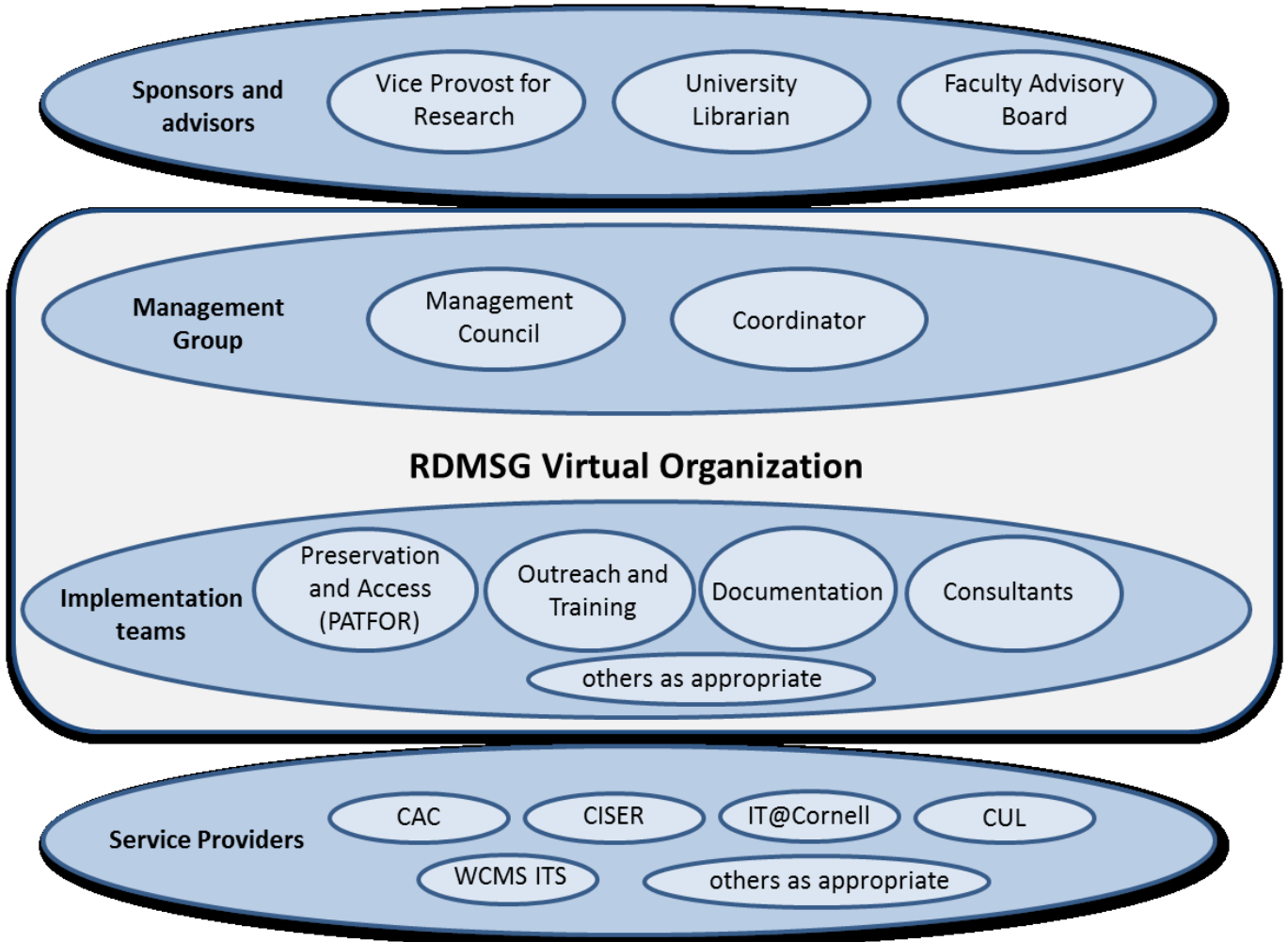
This mandate for increased public access to publicly funded research is another step in an apparent transition toward increased sharing and re-use of data. Not only do the agencies themselves need to plan for the impacts of open access, so too must academic institutions and the researchers themselves. With continued support and sponsorship by the University Librarian and the Office of the Vice Provost for Research, guidance and input by the Faculty Advisory Board and Management Council, participation of all related service providers and extensive expertise of the consultants, the RDMSG is well positioned to guide Cornell University forward in this expanding and exciting area of research support.

Based on feedback received from the RDMSG survey, next steps through 2013-2014 will include:

- Sharing of Data Management Plan examples as guidance for development of NSF DMPs.
- Dissemination of information regarding agency responses to the OSTP memo (above), including NSF and NIH, for increasing access to the results of federally funded scientific research.
- Broadened efforts for outreach and communication about RDMSG services.

¹ From the White House press release: <http://www.whitehouse.gov/blog/2013/02/22/expanding-public-access-results-federally-funded-research>; accessed 11/11/13.

Appendix A: RDMSG Organizational Structure



Appendix B: RDMSG Faculty Advisory Board Members

John Abowd (ILRLE)
Matt Aldridge (ORIA)
William Arms (CS)
Richard Burkhauser (PAM)
Claire Cardie (CS)
James Cordes (ASTRO)
Art DeGaetano (EAS)
Tom Frank (OSP)
Lawrence Gibbons (PHYS)
Natalie Mahowald (EAS)
Michael Webster (Lab of O)
Wendy Kozlowski (ex officio, RDMSG Coordinator)

Appendix C: RDMSG Management Council Members

William Block (CISER)
James Cordes (ASTRO)
Dean Krafft (CUL-IT)
David Lifka (CAC)
Janet McCue (CUL)
Vinay Varughese (WCMC-IT)
David Vernon (CIT)
Wendy Kozlowski (ex officio, RDMSG Coordinator)

Appendix D: Full Text of June 2013 RDMSG Researcher Survey

Thank you for taking the time to participate in this brief survey, presented by the Research Data Management Service Group (RDMSG). We are looking for feedback about your experiences with proposal-related Data Management Plans, and about the support offered by the RDMSG at Cornell.

All survey questions are multiple choice, with the opportunity in a few places to provide us with additional short answers. Questions marked with an asterisk (*) require an answer to move forward. The entire survey should take no longer than five minutes to complete.

If you have any questions or would like to contact us for any reason, please feel free to email us at rdmsg-help@cornell.edu. More information about data management services at Cornell is also available at our website, <http://data.research.cornell.edu>.

* If you submitted a proposal (or proposals) which required completion of a data management plan (DMP), did you receive specific feedback from reviewers or program managers with respect to any proposed DMP(s)?

- Yes
 No

Funding agency you submitted to that required a data management plan (select all that apply):

- | | |
|---|---|
| <input type="checkbox"/> National Science Foundation (NSF) | <input type="checkbox"/> National Institutes of Health (NIH) |
| <input type="checkbox"/> National Aeronautics and Space Administration (NASA) | <input type="checkbox"/> National Oceanic and Atmospheric Administration (NOAA) |
| <input type="checkbox"/> National Endowment for the Humanities (NEH) | <input type="checkbox"/> Institute of Museum and Library Services (IMLS) |
| <input type="checkbox"/> Other <input type="text"/> | |

NSF Office or Directorate, if appropriate (select all that apply):

- | | |
|--|---|
| <input type="checkbox"/> Office of the Director (Including cross-cutting or directorate-wide programs like GRFP, LTER, MRI, OIA, OISE, OPP etc.) | <input type="checkbox"/> Geosciences (GEO) |
| <input type="checkbox"/> Biological Sciences (BIO) | <input type="checkbox"/> Math & Physical Sciences (MPS) |
| <input type="checkbox"/> Computer & Information Science & Engineering (CISE) | <input type="checkbox"/> Social, Behavioral & Economic Sciences (SBE) |
| <input type="checkbox"/> Education and Human Resources (EHR) | <input type="checkbox"/> Environmental Research & Education (ERE) |
| <input type="checkbox"/> Engineering (ENG) | <input type="checkbox"/> Other <input type="text"/> |

We would appreciate hearing what feedback you received about your data management plan(s). Please use the space below to share input you've received on any of your proposal-related data management plans.

If you prefer to talk to us in person, enter your contact information here and someone from the Research Data Management Service Group will get in touch with you.

In the data management plans you've submitted, have you cited any of the following Cornell resources (select all that apply)?

- Cornell Center for Advanced Computing (CAC)
- Cornell University Library (including eCommons)
- Cornell Institute for Social and Economic Research (CISER)
- IT@Cornell
- Cornell Survey Research Institute (SRI)
- Computational Biology Service Unit (CBSU)
- Cornell Center for Technology, Enterprise and Commercialization (CCTEC)
- Other (please enter resource name here)
- None

* How have you, or anyone in your lab or group, used the Research Data Management Service Group (RDMSG) as a resource for data management information (select all that apply)?

- email consultation
- phone consultation
- in-person consultation
- visited website
- attended information session on NSF data management requirements
- other
- have not used RDMSG's services

How satisfied were you with your RDMSG interaction?

	Very Satisfied	Satisfied	Neither Satisfied nor Dissatisfied	Dissatisfied	Very Dissatisfied
Knowledge of relevant material	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professionalism	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Promptness of response	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

What other resources have you used for data management plan preparation?

- colleagues
- discipline-specific centers (ie. ICPSR, CIESIN, ORNL DAAC etc.)
- Cornell's Office of Sponsored Programs (OSP)
- other institutions' websites
- documents/policies from NSF directorate(s)
- direct contact with NSF program officer
- other
- none

Please use the space below for any suggestions of data management services that Cornell might expand or improve upon.

End of survey. If needed, please use the BACK button to edit any of your responses.

If you're satisfied with your responses, please press NEXT to submit.
