Cornell University's

Research Data Management Service Group

Summary of Activities and Accomplishments

Dec 2010 – June 2012

Contents

Introduction 1
RDMSG Outreach and Training Team 2
RDMSG Consultation Activities 4
RDMSG Documentation Team 6
Summary and Conclusions 9
References 10
Appendix A (Organizational Structure): 11
Appendix B (Faculty Advisory Board Members): 12
Appendix C (Management Council Members): 12

Introduction

In early 2008, grassroots discussions began among the Cornell research community aimed at addressing current and future challenges associated with data discovery, re-use and curation. Recognizing the need for a broad collaboration that bridges multiple scientific domains, library and archival science, computer and information science, the Data Intensive Science Organization for Virtual Exploration and Research (DISCOVER) team responded collectively to a call for proposals sponsored by the Vice Provost for Research. The team's proposal was successful and was funded by the Vice Provost's office in July 2008. As a result, the DISCOVER Research Service Group (DRSG) was formed as a partnership between domain scientists, the Center for Advanced Computing (CAC), the Cornell University Library, and Fedora Commons.

The Research Data Management Services Group, RDMSG, grew out of this initial collaboration. 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 (CIT), 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 inception through June 2012. 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 conducted 12 Data Management Plan Information Sessions, with over 300 total attendees.

The RDMSG established and engaged a pool of 12 consultants, with expertise ranging from engineering, math and science, to digital projects and intellectual property.

Consultants held 72 individual consultations, assisting faculty and student with Data Management Plans for NSF proposals and other data management needs.

The documentation team completely re-designed and updated the RDMSG website, and migrated it to a new platform.

The RDMSG website (data.research.cornell.edu) had over 8,100 visits, with over 3,000 of those visitors being from Ithaca (38%).

RDMSG Outreach and Training Team

Team members: Jeremy Cusker (Mar 2011 – Feb 2012), Wendy Kozlowski (Dec 2011 – present), Gail Steinhart (lead; Dec 2010 - present), Sarah Wright (Apr 2011 – present)

Charge: The RDMSG 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 RDMSG providers. The Outreach & Training team also has primary responsibility for marketing RDMSG 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 RDMSG Outreach & Training team is not charged with training RDMSG participants (consultants, service providers) or with handling communication tasks of the RDMSG such as ongoing communication with (or recruitment of new) RDMSG service providers.

Information Sessions on NSF Data Management Plans

Information sessions consumed most of the team's effort in its first 19 months. We held nine open sessions in 2011 (3 in January, 2 in May, 2 in October, and 2 in December), one session specifically for engineering faculty submitting CAREER proposals to NSF, and two sessions in May of 2012. Total attendance for all sessions was just over 300. Slides for all sessions (and video for one of the January sessions) are available on the RDMSG website: https://confluence.cornell.edu/display/rdmsgweb/nsf-data-management-plan-info-session-slides.

We solicited feedback after the December 2011 and May 2012 sessions, which was uniformly positive. When asked whether the session made them better prepared to address the NSF's data management plan requirement, 30 of 32 respondents answered "yes" and 2 answered "unsure." Here are some of the comments we received:

- "Website looks especially helpful including today's slides. Thank you!"
- "Down and to the point!"
- "Hearing that there are consults is very helpful."

Summary of Activities, Dec 2010 – Jun 2012

- Developed charge.
- Contributed content for the RDMSG website:
    - General data management planning guide: https://confluence.cornell.edu/display/rdmsgweb/data-management-planning-overview
    - Frequently asked questions (based on information sessions on NSF data management planning): https://confluence.cornell.edu/display/rdmsgweb/nsf-data-management-plan-faq
    - Guide to preparing tabular data for description and archiving: https://confluence.cornell.edu/display/rdmsgweb/preparing-tabular-data-description-and-archiving
- Held 12 information sessions on data management plans for NSF proposals (details above).
- Held 4 additional specialized information sessions on data management planning and RDMSG services

    - Humanities PhD Immersion Program (1/17/12)
    - Anthropology department faculty session (1/27/12)
    - Olin/Kroch/Uris grants workshop (4/5/12)
    - Molecular Biology and Genetics faculty session (4/6/12)
- Presented to various CUL groups to explain the new NSF requirement and promote the RDMSG: Reference & Outreach forum, Science Team, Mann Reference staff.
- Met with CUL communications group to develop a marketing and communications plan.
- Worked with CUL communications to develop graphic identity for RDMSG, including bookmarks for distribution/outreach.
- Established Twitter account to use for marketing; Twitter feed is embedded on the RDMSG website.
- Added RDMSG to VIVO.
- Developed a glossary of data management terms for the RDMSG website.
- Requested and received links to RDMSG website from CAC, CISER and OSP.

Moving Forward:

- Continue with NSF information sessions (October/fall break, December/study period) as long as attendance warrants this.
- Continue to develop content for the RDMSG website, as needed. Currently planned: best practices for readme-style metadata files.
- Develop a "We recommend" for the CUL website (replace the old one about the old research data management services directory).
- Develop slides for CUL LCD screens for general RDMSG promotion.
- Complete inventory of new faculty orientation events and try to participate whenever possible.
- Consider offering an information session targeted at graduate students working on dissertation improvement grants, or other sessions tailored specifically for graduate students.
- Solicit input at October 2012 information sessions to determine interest in other types of workshops, such as publishing data sets to eCommons, more domain focused information, etc.
- Continue to communicate with CUL liaisons to encourage referrals to RDMSG.
- Monitor emerging data management policies of funders such as NIH, and plan additional outreach accordingly (information sessions, Chronicle article).
- Other outreach opportunities as they arise.Explore with OSP opportunities for outreach and interaction with researchers.

RDMSG Consultation Activities

Consultant Team Members: Florio Arguillas (May 2012 – present), Aaron Birkland (Dec 2011 – present), Adam Brazier (Dec 2010 – present) ,Eric Chen (Dec 2010 – Jun 2011) Jeremy Cusker (Mar 2011 – Feb 2012), Dianne Dietrich (Dec 2010 – present), Peter Hirtle (Mar 2011 – present), Keith Jenkins (Oct 2011 – present), Wendy Kozlowski (coordinator; Dec 2011 - present), Stefan Kramer (Dec 2010 – Jan 2012), Wyman Miles (Oct 2011 – present), Fiona Patrick (Feb 2011 – present), Gail Steinhart (Dec 2010 – present), Linda Woodard (Dec 2010 – Dec 2011), Drew Wright (Dec 2011 – present), Sarah Wright (Apr 2011 – present)

Operating Principles:

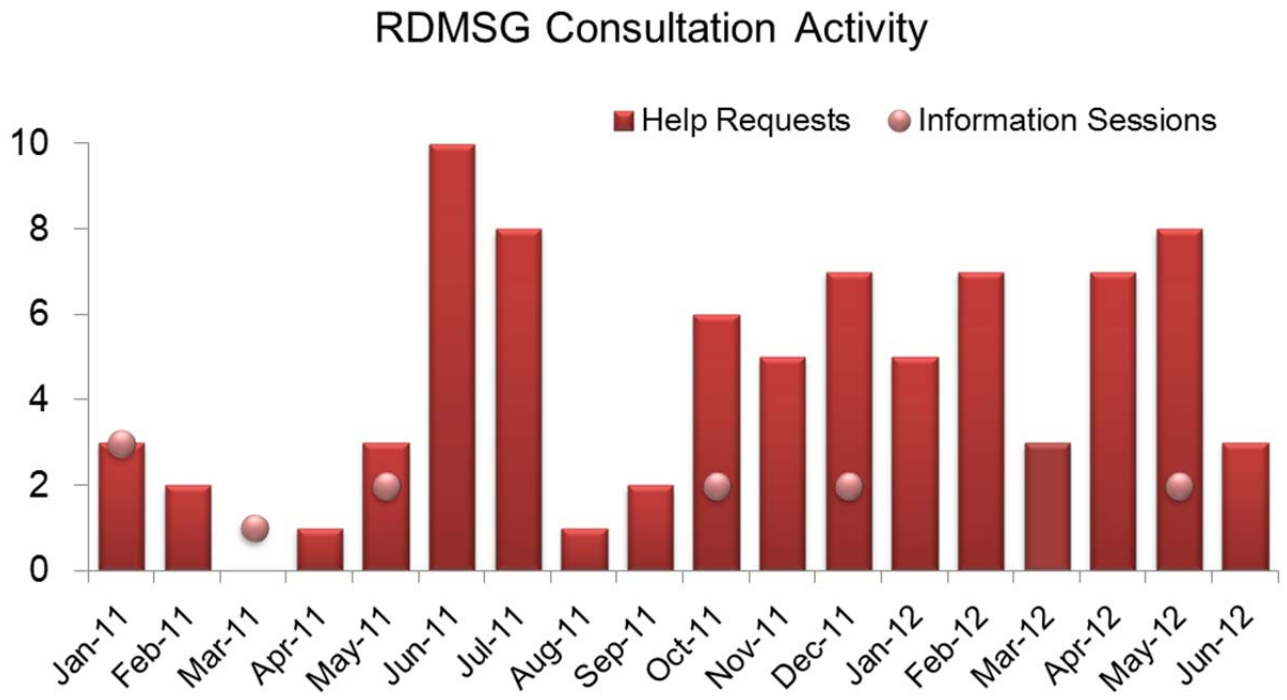
RDMSG consultants strive to meet the Cornell research community's data management needs by:

- 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 RDMSG and recognizing when additional expertise may be required to meet a researcher's needs.
- Sharing information with other RDMSG 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, Dec 2010 – June 2012

- Developed set of operating principles for guidance and direction of consultation activities
- Met regularly as a group to share experiences and information regarding consultation issues
- Held 72 individual consultations with Cornell faculty, staff and students.[1]
    - Nearly all consultations (~83%) involved questions directly related to NSF's Data Management Plan requirements
    - Average time spent on consultations was just over one hour (~70 minutes)
    - Most consultations were done via email (~72%), with the remainder being done face-to-face or by phone

[1] Summary information from these consultations reflects only those sessions with CU faculty, staff or students, where relevant data was recorded (e.g. consultation duration was not recorded for 6 of the 72 consultations; mode of communication was not recorded for 7 of the 72 consultations).



Moving Forward:

- Continue consultation efforts, for DMP-specific as well as related data management support needs.
- Hold regular meetings for active consultants to facilitate sharing of information and experiences.
    - Keep abreast of CU data management tools and services by inviting RDMSG service providers (e.g. CISER, CIT, CAC, CUL etc.) to attend monthly meetings.
- Maintain or expand the range of expertise and experience within the consultant pool by adding individuals as dictated by the needs of users.

RDMSG Documentation Team

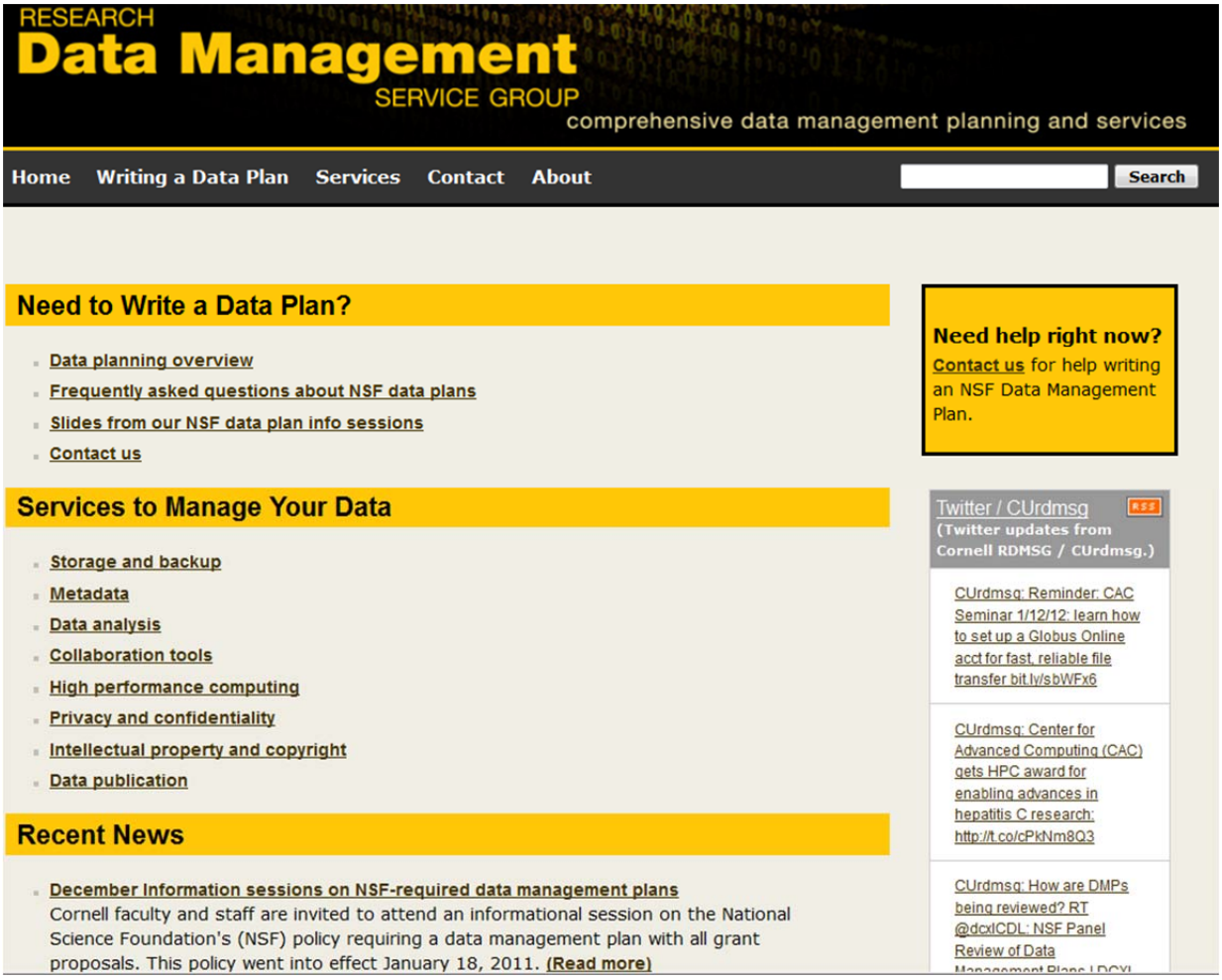
Team members: Eric Chen (Dec 2010 – Jun 2011), Dianne Dietrich (lead; Dec 2010 - present), Sarah Wright (Apr 2011 – present)

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.

RDMSG Website Overhaul

In 2011, the Documentation Team worked to move the existing RDMSG website from Drupal, hosted on a CAC server that was no longer being maintained, to Confluence. We used that move as an opportunity to improve the website's appearance, and developed a more polished website with the assistance of CUL's Communications team and a web designer. The resulting website is shown below, and can be accessed at http://data.research.cornell.edu.



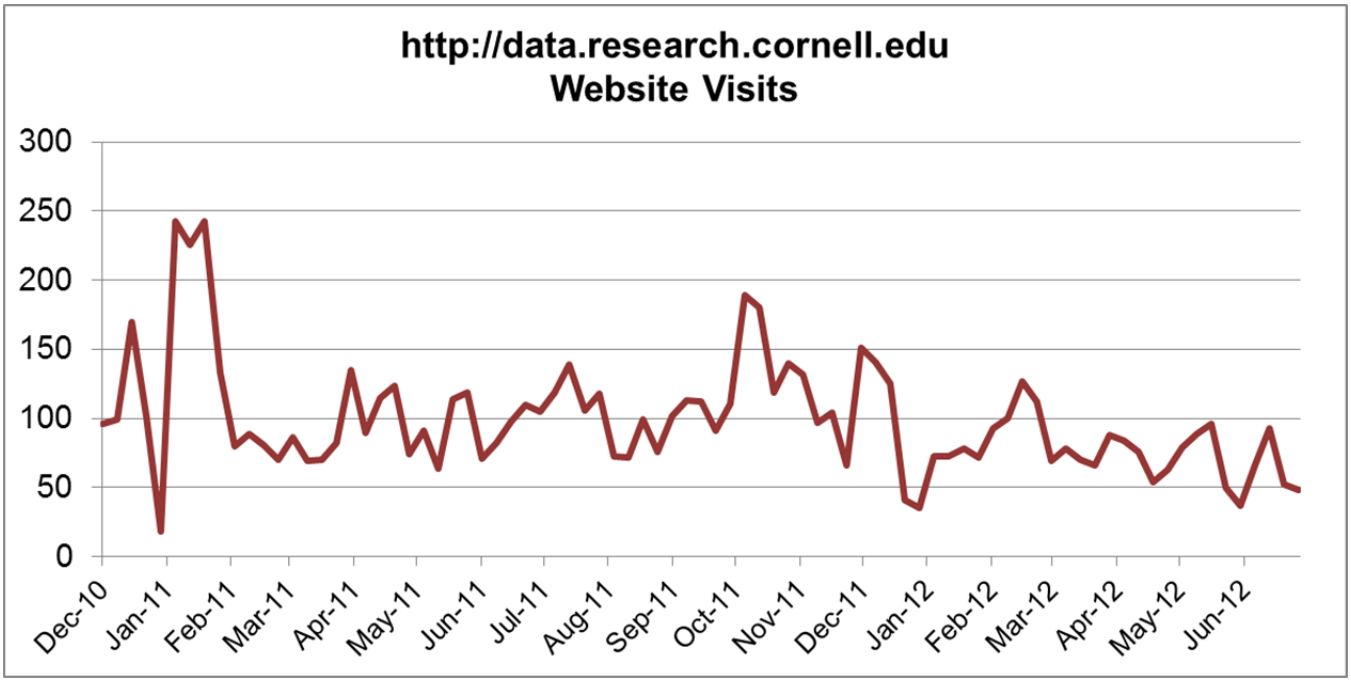
Summary of Activities, Dec 2010 – Jun 2012

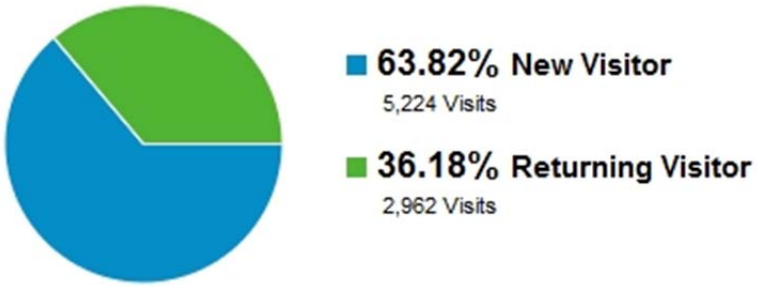
- Drafted charge and defined scope of activities

- 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.

- Secured long-term platform for RDMSG website
    - Investigated pros and cons of various hosting solutions for RDMSG Website
    - Migrated RDMSG website content from Drupal installation on CAC server to Confluence.
    - Developed new site design in consultation with a web developer

- Maintained RDMSG website
    - Served 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
    - Organized and updated content, in consultation with RDMSG Implementation Teams
    - Added information for researchers about funding requirements, service providers, the RDMSG team, and other information as appropriate.
        - General data management planning guide: https://confluence.cornell.edu/display/rdmsgweb/data-management-planning-overview
        - Frequently asked questions (Content from Outreach & Training Team, based on information sessions on NSF data management planning): https://confluence.cornell.edu/display/rdmsgweb/nsf-data-management-plan-faq
        - Guide to preparing tabular data for description and archiving: https://confluence.cornell.edu/display/rdmsgweb/preparing-tabular-data-description-and-archiving
        - Introduction to Intellectual Property Rights in Data Management (An overview of issues associated with managing intellectual property rights in data projects, prepared by Peter Hirtle, Senior Policy Advisor, Cornell University Library) : https://confluence.cornell.edu/display/rdmsgweb/introduction-intellectual-property-rights-data-management
        - About the consultants page with bios and link to VIVO profiles: https://confluence.cornell.edu/display/rdmsgweb/consultants





Challenges

- Securing web design assistance was very difficult. We needed both general design consultation and Confluence-specific implementation assistance. The solution ended up being simple, but determining which resources were available to us was difficult.

Moving Forward:

- Continue to develop content for RDMSG site
- Plan and implement refinements to site to improve usability and discoverability of content

Summary and Conclusions

In 2011, Cornell researchers submitted 465 proposals to the NSF; this represents nearly 16% of the total number of proposals and over $1.3 million dollars in research funding (nearly 20% of total proposed research in 2011; see also table 1 for additional information)[2]. Combining the efforts of the information sessions and the direct consultations, the RDMSG reached an estimated 24% of NSF submissions[3]. In 2012-2013, we hope to continue these efforts, and plan to reach out to CU researchers for feedback on the DMP review process, data management service needs and RDMSG-user satisfaction. This information will be anonymously shared via information sessions and consultations to help educate future proposers and the data will be used to shape RDMSG activities into the future.

In addition to the NSF policy, several other large funding agencies and publishers have data sharing requirements (e.g. NASA, NIH, Science, Nature) and there is every reason to believe this trend toward increased sharing, access and reuse of data will continue (Data and Informatics Working Group, 2012; Interagency Working Group on Digital Data, 2009; National Science Board, 2012). As time passes, it is expected that not only will the criteria for the DMP portion of such proposals be made more concrete[4] but related demands on data management services and infrastructure to fill the proposed plans for storage, sharing and archival of research data are also likely to increase.

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.

[2] Total number and value of NSF proposals submitted does not include those sub-contracted through other universities.

[3] Based on 90% of information session attendees being academic faculty, research staff or students, a conservative estimate of 25% of those attendees submitting a single NSF proposal in 2011, plus consultations that directly addressed the NSF DMP requirement.

[4] From keynote speech by Myron Gutmann, Assistant Director and Head of SBE Directorate, NSF, at the 3rd Annual Symposium for Open Access, held at the University of North Texas in Denton, TX; 21 May 2012.

Table 1: CU Research Proposal and Awards, 2011

	n	Total Value (in millions USD)
All Funders		
Proposals	2956	1367.1
Awards	1808	684.8
NSF		
Proposals	465	268.9
Awards	303	152.9

Source: Office of Sponsored Programs, Proposal and Award data for 2011.

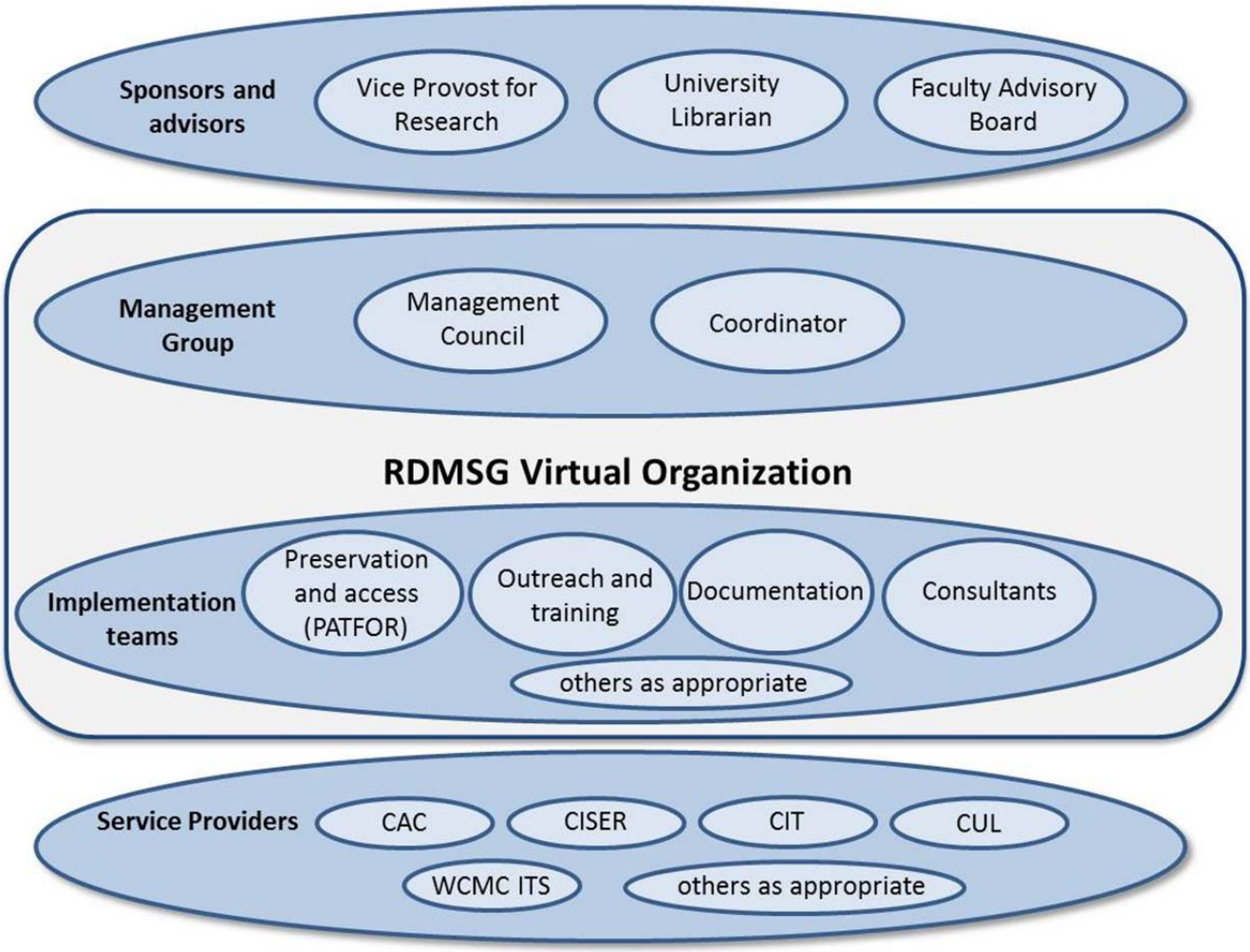
References

Data and Informatics Working Group. (2012). *Draft report to the advisory committee to the director.* Bethesda, MD: National Institutes of Health. http://acd.od.nih.gov/Data%20and%20Informatics%20Working%20Group%20Report.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.* Arlington, VA: Networking and Information Technology Research and Development (NITRD) Program. http://www.nitrd.gov/About/Harnessing_Power_Web.pdf

National Science Board. (2012). *Digital research data sharing and management.* (No. NSB-11-79). Arlington, VA: National Science Foundation. http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf

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)