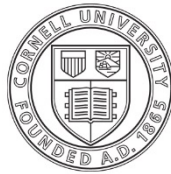Cornell University's

Research Data Management Service Group

Summary of Activities and Accomplishments
July 2013 – June 2014

Contents

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 2013 through June 2014. Two implementation teams, Outreach and Training (including the Website Maintenance subgroup) and Consultants were active in this time. 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 3 campus events, teaching 13 specialized data management-related courses or workshops, and holding 2 general DMP information sessions.
>
> Consultants received 56 requests for assistance, helping faculty, staff and students with their data management needs. Of those, 26 were inquiries directly related to proposal data management plans, and another 20 were requests for general data management assistance or information.
>
> 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.

RDMSG 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 RDMSG Outreach & Training (O&T) 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.

Overview of Major Activities

- The website was a primary focus for the team this year, and the new Drupal-based site went public May of 2014.

- The Outreach team held two formal information sessions this year, both targeting those who might be unfamiliar with RDMSG services but also covering basics of Data Management Plan preparation for grant proposals.

- In addition to the organized information sessions, the outreach team participated in a number specialized events where people with a particular need or of a certain user group would be more likely to attend. See the list of outreach and training activities that follows.

- We continued holding walk-in consultations held at various locations around campus (more information below).

Summary of Outreach and Training Activities

- Center for Vertebrate Genomics Annual Symposium (7/11/13): Wendy Kozlowski presented a poster. Attendance: not recorded.

- Inside Scoop (9/10/13): Participated in a library orientation event for graduate students. Olin Library. Talked directly to 35 students.

- Managing Data to Facilitate your Research: Introduction to Data Management for Chemistry Graduate Students (9/17/13). Presentation by Sarah Wright. Attendance: 6.

- Show Me the Money (10/22/13): Participated in graduate student workshop on obtaining research funding. Olin Library. Presentation by Sarah Wright. Attendance: 9.

- Immersion program for graduate students in the humanities (01/15/14): Taught a workshop on managing digital content, including humanities research data. Presented by Wendy Kozlowski. Attendance: 7.

- Managing Data to Facilitate your Research: Introduction to Data Management, for Chemistry and Physical Sciences students and staff (01/27/14): Presented by Sarah Wright. Attendance: 17.

- Horticulture Seminar (02/03/14): Introduction to RDMSG. Presented by Sarah Wright. Attendance: ~30.

- New Faculty Data Management Planning Information Session (1/16/14): Information session in Mann Library for new faculty indicating interest in RDMSG on survey from the Office of the Vice Provost for Research. Presented by Wendy Kozlowski Attendance: 17.

- New Faculty Data Management Planning Information Session (1/16/14): Information session in Clark Hall for new faculty indicating interest in RDMSG on survey from the Office of the Vice Provost for Research. Presented by Wendy Kozlowski Attendance: 7.

- 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.

- Intro to Excel Workshop (3/11/14): 1.5 hour workshop on best practices, tips and tricks for using Microsoft Excel to work with research data. Presented by Wendy Kozlowski and Sarah Wright. Attendance: 10.

- Excel – Formulas and Functions Workshop (3/18/14): 1.5 hour workshop on best practices, tips and tricks for using Microsoft Excel formulas and functions to work with research data. Presented by Wendy Kozlowski and Sarah Wright. Attendance: 16.

- Excel – Macros Workshop (4/08/14): 1.5 hour workshop on how-to's and best practices for simplifying data analysis using Microsoft Excel macros with research data. Presented by Dianne Dietrich and Wendy Kozlowski. Attendance: 16.

- Excel – Pivot Tables Workshop (4/15/14): 1.5 hour workshop on how-to's and best practices for simplifying data analysis using Microsoft Excel pivot tables with research data. Presented by Wendy Kozlowski. Attendance: 10.

- Organize, Safeguard, Update: Managing your Digital Research Files workshop (4/30/14): With Danielle Mericle, taught workshop on best practices for managing digital content, including an introduction to services available through RDMSG. Presented by Wendy Kozlowski. Attendance: 7.

- Software Bootcamp (6/5/14): Informal information session on RDMSG. Presented by Daina Bouquin. Attendance not recorded.

- IT@Cornell - Community Conference (6/18/14): Presented a poster at Cornell IT community event. Attendance: ~350 for entire event.

- Show Me the Money (6/25/14): Participated in graduate student workshop on obtaining research funding. Olin Library. Attendance: 1.

Walk-in Consultations: In an effort to keep RDMSG in the public eye and to make our services more visible, we've continued to offer walk-in consultations (http://data.research.cornell.edu/calendar/ lists schedule and locations). We advertise these via LCD displays located in Mann Library, Duffield atrium, Physical Sciences, Space Sciences, Food Science, and the College of Veterinary Medicine. 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.

Despite limited uptake, overhead is low for this effort and we have continued offering walk-in hours throughout the year. Discontinuing them would mean taking down the LCD displays, and we speculate that this is good advertising, and if taken down, 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, participating in OSP roundtables to inform research administrators of data management issues and the services RDMSG can provide, and presentations and publications beyond Cornell.

Plans for 2014-2015

- Continue to develop content for the RDMSG website, as needed. Some topics being considered for development are best practices for data deposit to eCommons, data preservation and archiving and best practices for ELN use.
- 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.
- Engage in other outreach opportunities as they arise. Explore with OSP, ORIA, BRC, CVM and other campus groups opportunities for interaction with faculty, staff and research administrators.

RDMSG Website Update

As part of the Outreach and Training activities, the RDMSG website was overhauled and updated, and the revamped website went public in May of 2014. The new design allows for broader editing capabilities, allowing what was formerly the Documentation team to share more broadly among O&T team members the responsibility for site maintenance. As such the "Documentation Team" has become the "Website Maintenance" sub-team under Outreach and Training, and Dianne Dietrich and Sarah Wright continue to take primary responsibility for updates, communication with Drupal designers, and scheduling of content updates with service providers.

The RDMSG website includes 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 Outreach and Training team work with any relevant Cornell-related groups, including other RDMSG implementation teams and web design teams, all as appropriate.

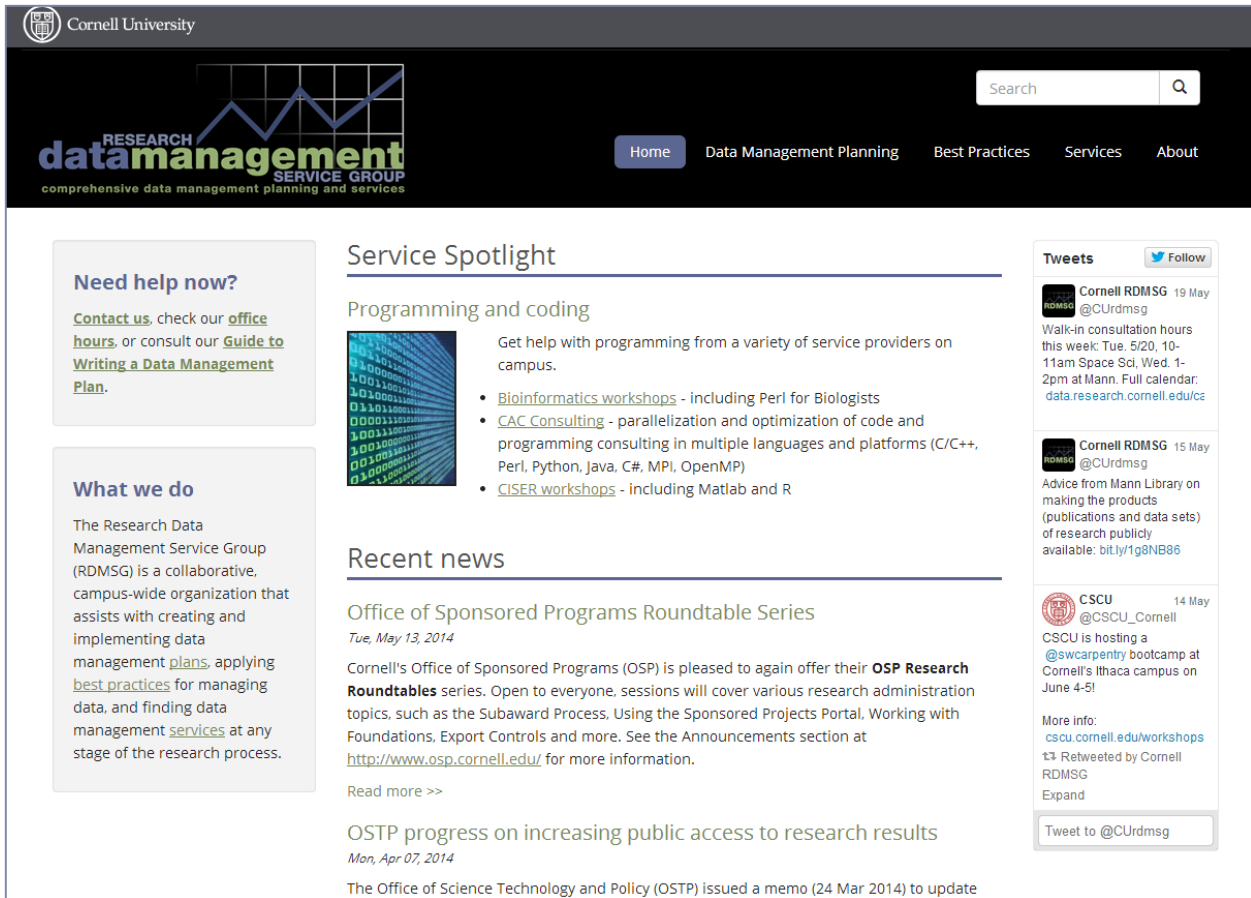*Figure 1: RDMSG website before redesign, using Confluence.*

Figure 2: RDMSG Website after redesign, using Drupal

Documentation Team Activities

- Redesign of RDMSG Website
  - CUL Usability walkthrough of website suggested it had outgrown its old structure because of new content types being added.
  - Usability methods were employed by the RDMSG Outreach & Training team to re-think website structure and organization.
  - As a result, separated Cornell data management services from best practices guidance, creating separate sections for each on the website.
    - Created new best practices guidance:
      - Data Citation
      - File Formats
      - File Management
      - Guide to Writing a Readme File
      - Metadata and Describing Data
      - Sharing Data
    - Significant revision of existing guidance documents:
      - Guide to Writing a Data Management Plan (DMP)
      - Introduction to Intellectual Property Rights in Data Management
    - Note that documents were authored by all members of the Outreach & Training team.
  - Worked with relevant CUL support staff to develop a new logo and website design.
- Implemented changes for redesigned RDMSG Website

- o   Moved from Confluence to Drupal, necessitating planning for content types and overall site architecture.
- Maintained RDMSG website
  - o   Continued to serve as contact for updates, edits, and any other changes to the site.
  - o   Coordinated review of Service Provider links; updated links when needed.
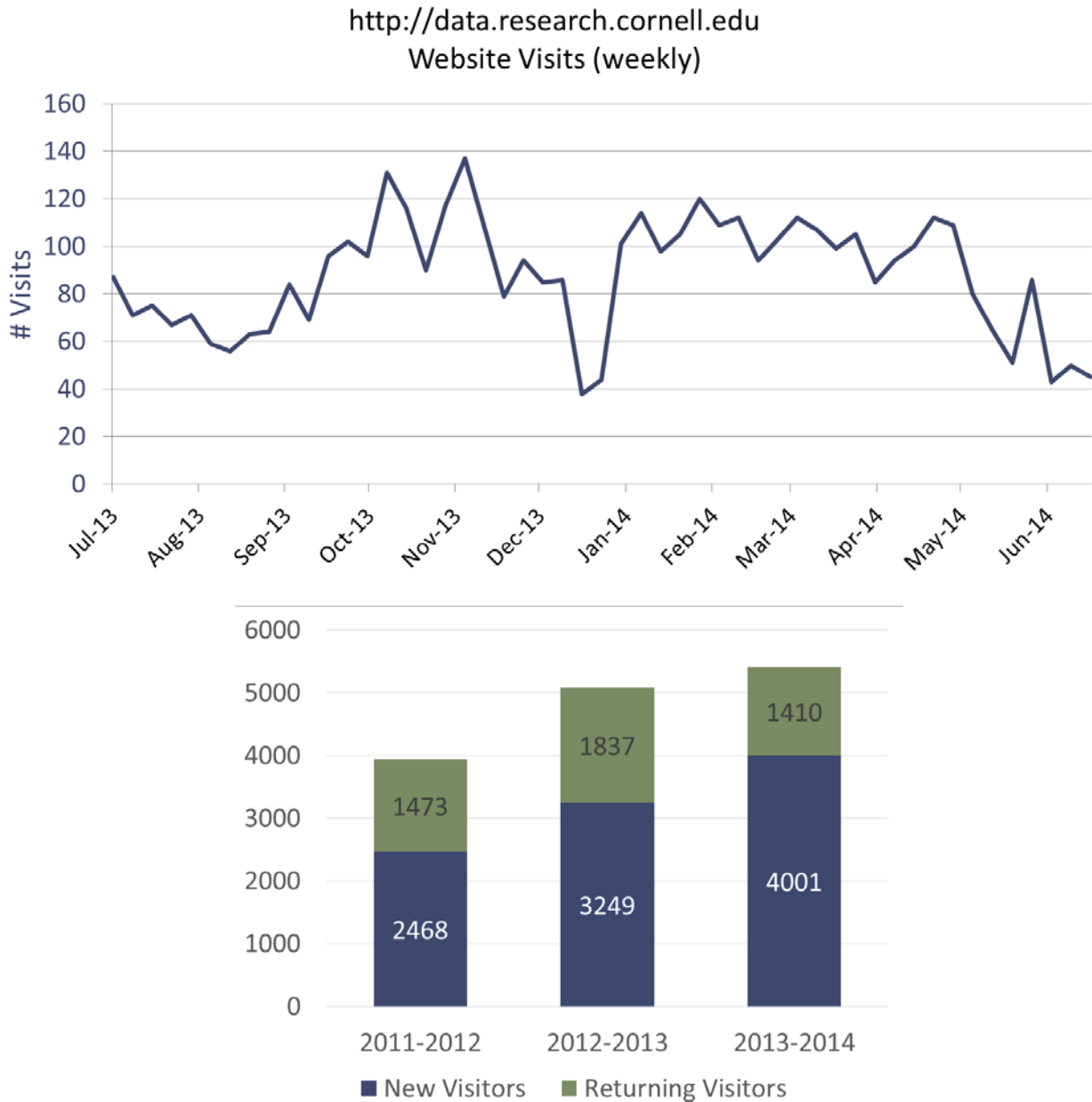

Figure 3: http://data.reserach.cornell.edu Usage statistics.

Plans for 2014-2015

- Will continue to create "how to" documents for maintaining the website.
- Continue to maintain RDMSG site by monitoring external links and implementing refinements to improve usability and discoverability of content.

RDMSG 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:

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, July 2013 – June 2014

- Met regularly to share experiences and information regarding consultation issues, as well as developments at Cornell and elsewhere related to data management.
- Customized the updated version of the Data Management Plan-preparation tool (DMPTool: http://dmptool.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.
    - A link to the DMPTool (http://data.research.cornell.edu/content/data-management-plan-tool) can be found on the Data Management Planning page of the RDMSG website, and the tool is additionally featured as one of the "Service Spotlights" on the home page.
    - While there are still no Cornell DMPs approved for sharing publicly, the DMPTool provides some publicly accessible sample DMPs (https://dmptool.org/public_dmps) and additional samples can be found on the DMPTool templates page (https://dmptool.org/guidance). Like those found elsewhere online (such as http://idi.ucsd.edu/data-curation/examples.html, http://www.irss.unc.edu/odum/contentSubpage.jsp?nodeid=570 and others), it is important that researchers realize these are examples only and it is critical that their own proposed data management plans be customized to their own systems, workflows, needs and norms of their discipline.
- We received 56 inquiries to RDMSG between July 2013 and June 2014. The majority came in via the RDMSG RT ticketing system (rdmsg-help@cornell.edu); those that were directly addressed to

specific consultants were later added to RT for record-keeping. In total, these inquiries led to 46 individual consultations with Cornell faculty, staff and students.

- While many consultations still are inquiries directly related to preparation of data management plans (46% of the total, with 37% NSF, 9% other funders), this year the number of general data management inquiries rose (36% of the total, as compared to 19% in 2012-2013). The remaining 18% of the inquiries to the ticketing system were administrative in nature.
- Average time spent on consultations remains close to one hour (~50 minutes).
- The percentage of face-to-face consultations increased relative to the last two years, with over half (28 of 46) of the consultations this year being done in person.
- Non-DMP-related consultations included conversations with institutions outside of Cornell regarding organization and logistics for data management support and outreach, metadata creation inquiries, database design and support questions, follow-ups on execution of a DMP, copyright issues related to data sharing and requests for data analysis support in Excel.
- In addition to the above numbers, there were at least 17 electronic lab notebook (ELN) related consultations, trainings and outreach sessions where the focus was use ELNs and the LabArchives product. Interested groups were from both in and outside Cornell University (eg. Columbia, Stanford, Pitt, Tufts and others). For these consultations, Wendy Kozlowski represented RDMSG to support research lab use of ELNs.
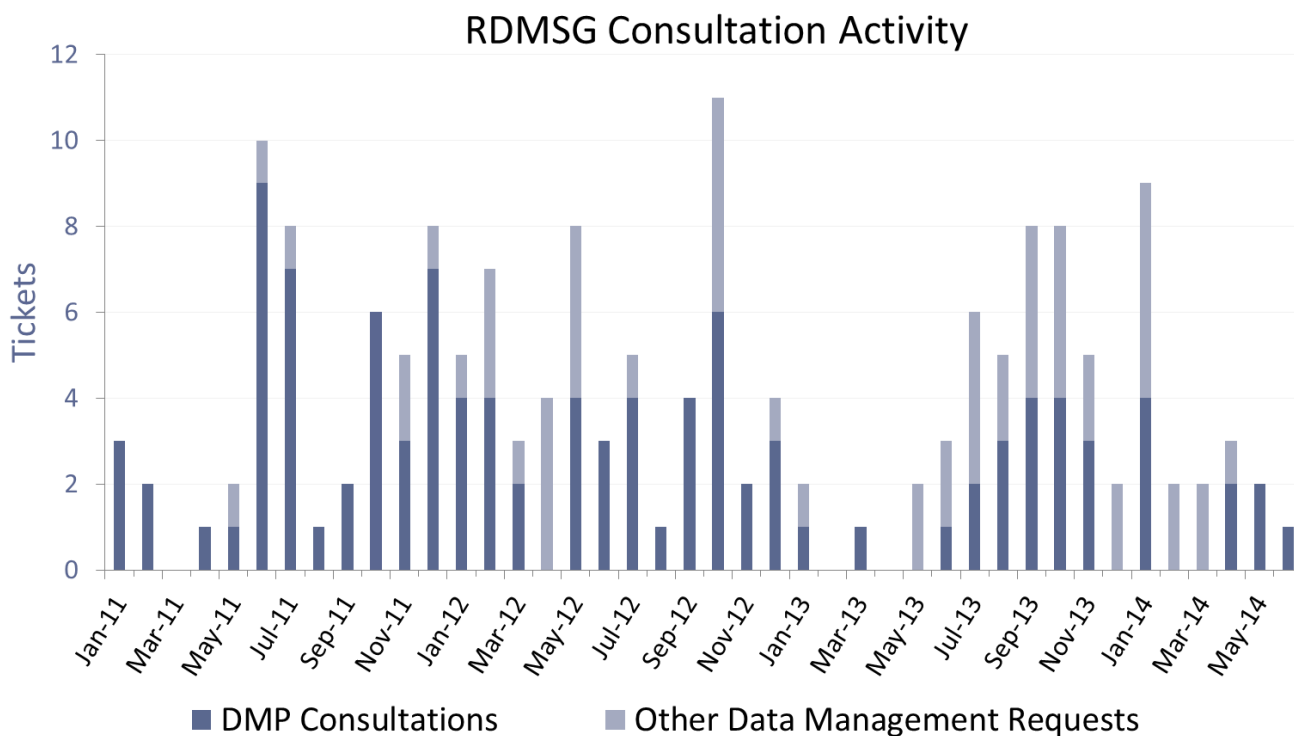

Figure 4: RDMSG Consultation activity, 2011 – 2014.

<u>Plans for 2014-2015</u>

- 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.

RDMSG Publications and Presentations

Kozlowski, Wendy. 2014. Building Collaborative Data Management Support at Cornell University. Doing it *Your* Way: Approaches to Research Data Management for Libraries, National Network of Libraries of Medicine Symposium. April 28, 2014. The Rockerfeller University, New York, NY. http://nnlm.gov/mar/libraries/data2014/kozlowski2014.pdf

Steinhart, Gail. 2014. An Institutional Perspective on Data Curation Services: A View from Cornell University. In: Managing Research Data: Practical Strategies for Information Professionals. Joyce Ray, editor. Purdue University Press, West Lafayette, IN.

Summary and Conclusions

As far back as 1995, agencies such as the National Science Foundation (NSF) have had statements in their grant and award proposal preparation guides (http://www.nsf.gov/bfa/dias/policy/dmp.jsp) stating the expectation of sharinging research results. More recently, agencies such as the NIH (2003) and the NSF (2011) have instituted requirements for proposals to include a written plan that describes researchers' intentions for data management and sharing.

With the 2013 memo from the White House's Office of Science, Technology and Policy director that directed "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", (http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf), more and more agencies are following along a similar DMP requirement path.
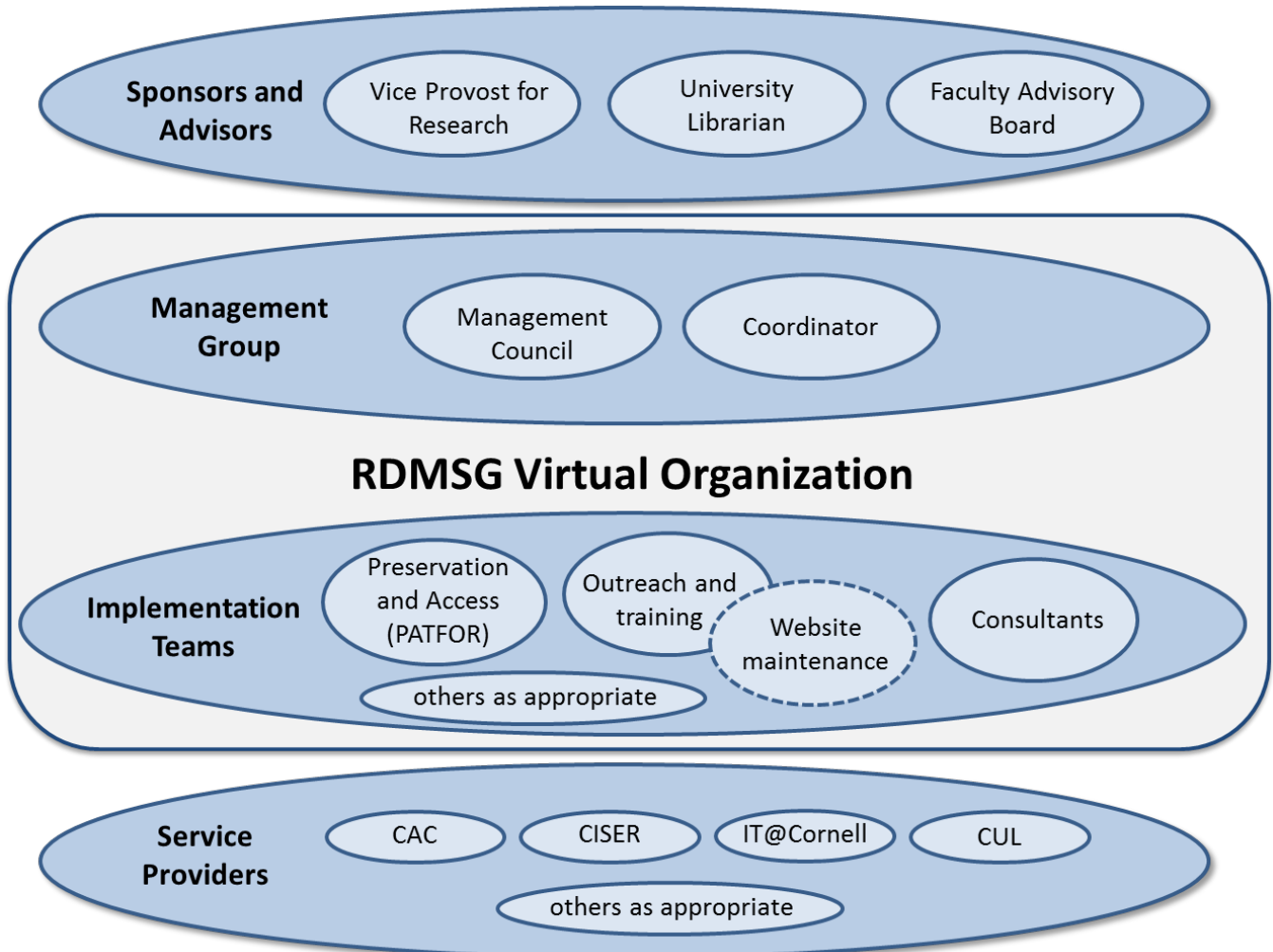
At Cornell, many of the consultation requests that come to the RDMSG are to help create or review such Data Management Plans (DMPs). The majority of those are for proposals being submitted to the NSF, but we also see requests for help with agencies such as the NIH, NASA and NOAA. As more and more agencies enact such planning and sharing requirements, the RDMSG consultants will continue to familiarize themselves with agency regulations to be able to assist researchers in proposal preparation.

In addition, the RDMSG consultants are beginning to receive requests for help from researchers after grants have been funded and data management plans need to be enacted. This interaction offers additional insights to what researchers at Cornell need to carry out their actual data management tasks. Whether issues revolve around workflow, storage, sharing or archiving, these consultations can provide practical insights to the types of tools and services needed to facilitate and simplify this portion of the scientific research workflow.

During 2014-2015, the RDMSG will focus efforts on:
- Dissemination of information regarding agency responses to the OSTP memo (above), for increasing access to the results of federally funded scientific research.
- Gathering information on researcher workflows and needs for enactment of proposed data management plans.
- Broadened efforts for outreach and communication about RDMSG services.
- Adding and updating relevant content to the RDMSG website for use by Cornell faculty, staff and students.

Appendix A:
RDMSG Organizational Structure

Sponsors and Advisors
- Vice Provost for Research
- University Librarian
- Faculty Advisory Board

Management Group
- Management Council
- Coordinator

RDMSG Virtual Organization

Implementation Teams
- Preservation and Access (PATFOR)
- Outreach and training
- Website maintenance
- Consultants
- others as appropriate

Service Providers
- CAC
- CISER
- IT@Cornell
- CUL
- others as appropriate

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)
Oya Rieger (CUL)
Vinay Varughese (WCMC-IT)
David Vernon (CIT)
Wendy Kozlowski (ex officio, RDMSG Coordinator)