Cornell University's

Research Data Management Service Group

Summary of Activities and Accomplishments
July 2019 – June 2020

Contents

Cornell University

RESEARCH DATA MANAGEMENT SERVICE GROUP

Comprehensive Data Management Planning & Services

Find solutions to store and share data.
Learn about best practices.
Write successful data management plans.
Consult with experts.

✉ 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 and implement 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), the Cornell Biotechnology Research Center (BRC), 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 2019 through June 2020. Both the Outreach and Training (including efforts to maintain and update the RDMSG website) and Consultants implementation groups were active in this time. This report includes summarizes their accomplishments, including the following highlights:

> The Outreach and Training team held a variety of workshops and information sessions, introducing faculty, staff and students to the RDMSG at 8 campus events, teaching in 3 departmental courses or seminars, and holding 11 specialized data management-related classes or workshops.
>
> RDMSG Consultants received 143 requests for assistance, resulting in 100 consultations where we helped faculty, staff and students with their data management needs. Of those consultations, 23 were inquiries directly related to proposal data management or sharing plans, and another 77 were requests for general data management assistance or information. The consultant pool continues to shift and grow as individuals come and go from Cornell, but currently includes 22 individuals with a wide variety of data expertise.
>
> The RDMSG, with financial support from the Cornell University Library, hosted a full-day data conference to bring together researchers – faculty, staff and students no matter their experience or field of work – to share and learn about working with data. The 2019 Day of Data featured a keynote lecture by Professor David Mimno, a morning panel and discussion, lunch with a data services "fair", and multiple tracks of afternoon workshops. With nearly 80 in-person attendees, up to 64 registrants for the workshops, plus virtual attendees, the event was a success and is hopefully just the first of what will become regular Cornell Day of Data gatherings.

RDMSG Outreach and Training Activities

Team members: Florio Arguillas (June 2018 – present), Erica Johns (Sep 2014 – present), Amelia Kallaher (March 2019 – present), Lynda Kellam (August 2019 – present), Wendy Kozlowski (Dec 2011 – present), Zoe Saldana (June 2018 – September 2019), Henrik Spoon (June 2018 – present), Sarah Wright (April 2011 – present, chair Jan 2018 – present).

Charge: The RDMSG Outreach and Training (O&T) team develops and delivers training materials for Cornell researchers on issues of data management, with particular attention to meeting the requirements of research funders, and to services available through RDMSG providers. The Outreach and Training team also has primary responsibility for marketing RDMSG services to the Cornell community, and 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, including faculty, students and staff. The RDMSG Outreach and 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

Reviewed and updated funder data requirements content on website, adopting a new template that is clearer and more useful for faculty new to creating data management plans. Major sections of the template are action-based, and now include:

- What is required?
- What do I need to submit as part of my proposal?
  - What is a data management plan?
- When do I need to share my data?
- Where can I deposit my data?
- How do I submit?
- Are there any compliance requirements?

Continued to build relationship with CUL liaisons to encourage referrals to RDMSG. The Outreach and Training team held the workshop "Data Engagement Opportunities: applying what you already know as a liaison" as an introductory workshop for liaison librarians (August 2019).

Hosted the first annual Cornell Day of Data on October 15, 2019. Offered as a full-day event at the ILR Conference Center, the *Day of Data 2019: Any person. Any study. Any data.* included a keynote address and discussion panel in the morning, a data services fair to highlight campus data services to help make their data Findable, Accessible, Interoperable, and Reusable (FAIR) during lunch, and two tracks of afternoon sessions and workshops. The event drew nearly 80 in-person attendees, as many as 64 registrants for the workshops, and many more virtual attendees. David Mimno from the department of Information Science at Cornell University, gave the keynote: "Data Science and Data Scholarship".

The Outreach and Training team participated in 23 specialized events (e.g. workshops, information fairs, learning groups) attended by approximately 360 people with a particular need or of a certain user group. Despite the move to a virtual campus in Spring of 2020, these teaching sessions continued throughout the

academic session. Table 1 details these activities. We also conducted a complete update of all of the Best Practices content on the website, nine comprehensive guides on data management topics such as Data Sharing, Data Citation, Guide to Writing "Readme" Style Metadata, Preparing Tabular Data for Description and Archiving, etc.

Date	Subject	RDSMG Personnel	Attendance	Notes
15 Aug 2019	Data Engagement Opportunities Workshop for CUL, specifically liaison librarians	Sarah J. Wright	11	Workshop
26 Aug 2019	New Graduate & Professional Student Orientation	Amelia Kallaher, Erica Johns	104	Campus Information Session
12 Sep 2019	Inside Scoop	Wendy Kozlowski, Henrik Spoon	30	Campus Information Session
17 Sep 2019	Data Organization for Spreadsheets	Wendy Kozlowski, Amelia Kallaher	7	Workshop; 21 registrants - 4 attended plus 3 walk-ins
19 Sep 2019	RDMSG and ORCiD	Wendy Kozlowski, Sarah J. Wright	25	Campus Information Session
24 Sep 2019	Data Documentation	Wendy Kozlowski, Amelia Kallaher	3	Workshop
04 Oct 2019	RIT Publishing Forum	Wendy Kozlowski	13	Off-campus Symposium
08 Oct 2019	Chemical Engineering Graduate Student Association	Wendy Kozlowski	11	Workshop
08 Oct 2019	David Levitsky Lab	Wendy Kozlowski	30	Course-related instruction
11 Oct 2019	New Faculty OSP Orientation	Wendy Kozlowski	15	Campus Information Session
18 Oct 2019	Rachel Kerr Lab	Wendy Kozlowski	11	Workshop
22 Oct 2019	Excel for Research: Formulas and Functions	Wendy Kozlowski, Sarah J. Wright	5	Workshop; 24 registered, 3 on waitlist
29 Oct 2019	Excel for Research: Pivot Tables	Wendy Kozlowski, Sarah J. Wright	6	Workshop
12 Nov 2019	Excel for Research: Macros	Wendy Kozlowski, Sarah J. Wright	4	Workshop; 24 registered; 15 on waitlist
13 Jan 2020	Introduction to Data Visualization in RStudio	Amelia Kallaher, Sarah J. Wright	9	Workshop
15 Jan 2020	Taking good care of your data	Wendy Kozlowski	13	Workshop
13 Feb 2020	Love Your Data Pop-Up	Sarah J. Wright, Erica Johns	8	Campus Information Session
13 Feb 2020	Love Your Data Pop-Up	Sarah J. Wright, Lynn Thitchener	20	Campus Information Session
28 Feb 2020	DMP guidance for AAP Faculty	Wendy Kozlowski	17	Campus Information Session
28 Feb 2020	Virtual Fair for potential grad students	Lynda Kellam, Wendy Kozlowski		Campus Information Session
10 Apr 2020	Data Viz and Management Best Practices	Wendy Kozlowski	7	Course-related instruction; PAM 4950
15 Apr 2020	Data Management	Wendy Kozlowski	3	Course-related instruction; ANTHRO 6440
01 May 2020	Data Management for lab group	Sarah J. Wright	8	Workshop

Table 1: Summary of Outreach & Training activities, 2019-2020

Marketing and Communication: Marketing and communication efforts continue to include web-based communication (i.e. announcements via website news items and Twitter), fliers in the library, use of campus LCD screens, emails, and word of mouth. In order to reach the broadest audience, we have utilized a variety of marketing and communication styles: LCD signs to promote awareness of our services, communicating with library liaisons to promote RDMSG referrals within their departments, tables at graduate orientation events, participation in OSP roundtables to inform research administrators of data management issues and the services RDMSG can provide, and presentations and publications beyond Cornell. We have continued to leverage CUL Communications staff to develop new and exciting marketing materials, including an assortment of new hex stickers on the theme of "Love your data" (see Fig. 1).

Figure 1: Love Your Data hex stickers

RDMSG Website

RDMSG continues to maintain a website to share information and best practices related to data management. We strive to keep the website as up to date as possible, and the team monitors content regularly.

One frequently used best practice page is about creating readme files for research data. Locally, we see usage for datasets going into the CUL repository, eCommons. The guide and downloadable template has been cited on many other data management websites (Montana State University, University of Illinois, Harvard, The Graduate Institute Geneva, Georgia Southern University, University of Wisconsin, Madison, Utah State University, The University of Oklahoma, Queens University Library and others), and is

regularly downloaded from around the world (Table 2). This year, the plain text template document was updated to meet web accessibility standards (since many repositories allow direct viewing of documents).

	Download Count
2019	
Jul	34
Aug	49
Sep	29
Oct	159
Nov	141
Dec	70
2020	
Jan	82
Feb	103
Mar	102
Apr	118
May	149
Jun	113
Grand Total	**1149**

Table 2: Download statistics for the RDMSG-created Readme template, July 2019-June 2020.

Plans for 2020-2021

- Review and update Funder requirements content on website.
- Host a Cornell Day of Data in January 2021.

RDMSG Consultation Activities

Consultant Team Members: Florio Arguillas (May 2012 – present), Eliza Bettinger (June 2017 - present), Adam Brazier (Dec 2010 – present), Meryl Bursic (Jan 2019 – present), Dianne Dietrich (Dec 2010 – present), Amy Dygert (Feb 2017 – present), Bobby Edamala (Nov 2019 – present), Zachary Jacques (June 2017 – present), Keith Jenkins (Oct 2011 – present), Erica Johns (Sept 2014 - present), Lynda Kellam, (Aug 2019 – present), Susan Kendrick (Sept 2019 – present), Wendy Kozlowski (coordinator; Dec 2011 - present), Rich Knepper (Nov 2019 – present), Whitney Kramer (Nov 2019 – present), Leah McEwen(Apr 2019 – present), Michelle Paolillo (Mar 2013 – present), Henrik Spoon (Nov 2017 – present), Lynn Thitchener (Nov 2019 – present), James Vanee (Mar 2016 – present), Drew Wright (Dec 2011 – present), Sarah Wright (Apr 2011 – present).

Previous Team Members: Aaron Birkland (Dec 2011 – Oct 2015), Daina Bouquin (Jul 2013 – Sep 2015), Tarek Chams (Nov 2016 – May 2019), Michelle Edwards (Jan 2015 – Dec 2016), Chris Fournier (Mar 2015

– Oct 2016), Peter Hirtle (Mar 2011 – Oct 2015), Tom Horton (Feb 2019 – Nov 2019), Wyman Miles (Oct 2011 – Aug 2018), Jaron Porciello (Mar 2013 – Aug 2015), Zoë Saldaña (June 2018 – Sept 2019), Gail Steinhart (Dec 2010 – Mar 2015)

<u>Operating Principles</u>:

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.

<u>Summary of Activities, July 2019 – June 2020</u>

- Consultants met monthly to share experiences and information regarding consultation issues, as well as developments at Cornell and elsewhere related to data management.
- We received 143 total inquiries to RDMSG between July 2019 and June 2020 (Figure 2). Many inquiries continue to reach us via the RDMSG RT ticketing system (rdmsg-help@cornell.edu); in addition, inquiries directly addressed to specific consultants were later added to RT for record keeping. In total, these inquiries led to 98 individual consultations with Cornell faculty, staff and students. The remaining 45 inquiries (31%) were either administrative in nature or related to the Day of Data event.
    - This year, 23 (23.5%) of our consultations were about preparation of data management plans related to a proposal (Figure 3). While we continue to see requests for help for NSF DMPs, we also consulted on data management or data sharing plans for USDA, NIH, NOAA, EPA, and CDC grant proposals.
    - Average time spent on consultations (non-administrative) remains just under an hour (~50 minutes), though consultation times ranged from just 5 minutes to as long as 2-4 hours in some cases.
    - While the majority of consultations have historically been done over email, due of course to the onset of COVID19, nearly all our consultation after March 2020 were virtual, mostly over Zoom.
    - The largest numbers of consultations (where recorded) were from Cornell's College of Agriculture and Life Science, the College of Arts and Sciences and the College of Engineering (Figure 4)
    - Non-DMP-related consultations included conversations with institutions outside of Cornell regarding organization and logistics for data management support and outreach, and inquiries from within Cornell for metadata and documentation creation assistance, legal support inquiries, data storage advice, follow-ups on execution of DMPs, and requests for data analysis and cleanup support in Excel, Open Refine, SPSS, R and python.
- In addition to the above numbers:

o   There were at least 7 electronic lab notebook (ELN) related consultations, trainings and outreach sessions where the focus was use ELNs and the LabArchives product and/or the Open Science Framework product. Wendy Kozlowski continues to represent RDMSG to support research lab use of ELNs.
o   Wendy Kozlowski and Sarah Wright curated a total of 34 datasets for deposit into eCommons. Curation services for those datasets included format and content review, dataset documentation and discovery metadata creation assistance, and either submission on the student or faculty's behalf, or training on the eCommons system for user self-submission.

Monthly RDMSG Consultation Activity

DMP Consultations   **Other Data Management Consultations and Inquiries**
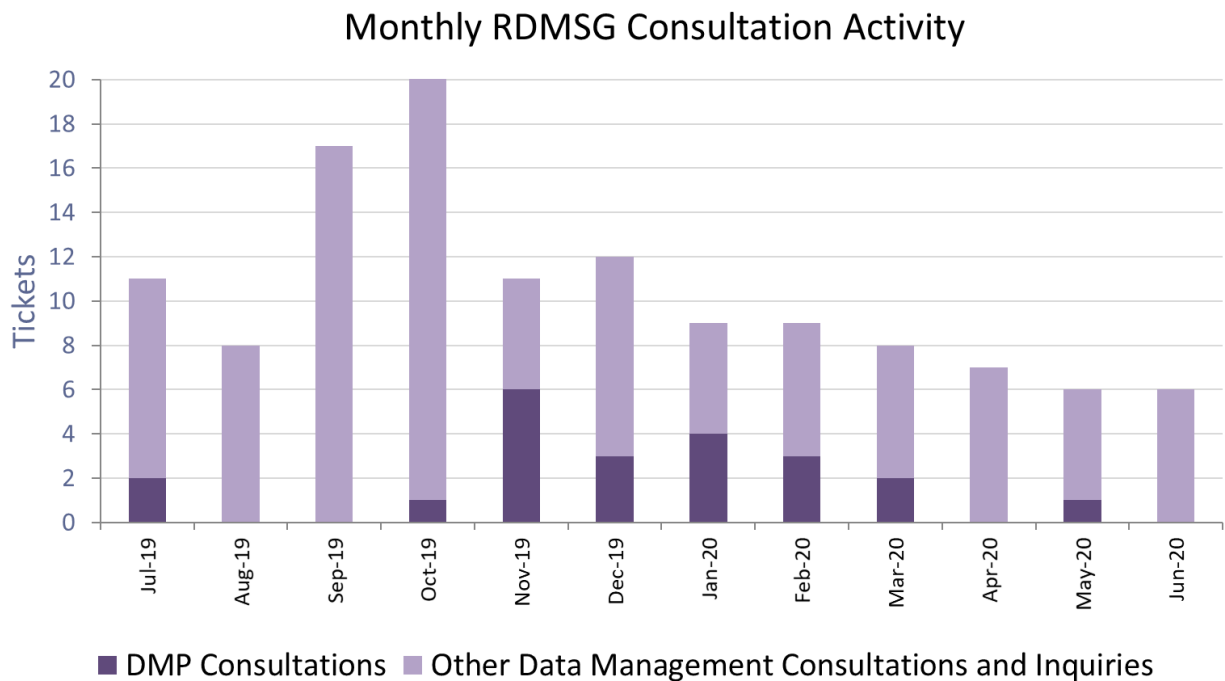
Figure 2: RDMSG Consultation activity, 2019 – 2020, showing an increase in activity around the Day of Data, and increased DMP consultations around the turn of the calendar year.

RDMSG Inquiries and Consultations by Type

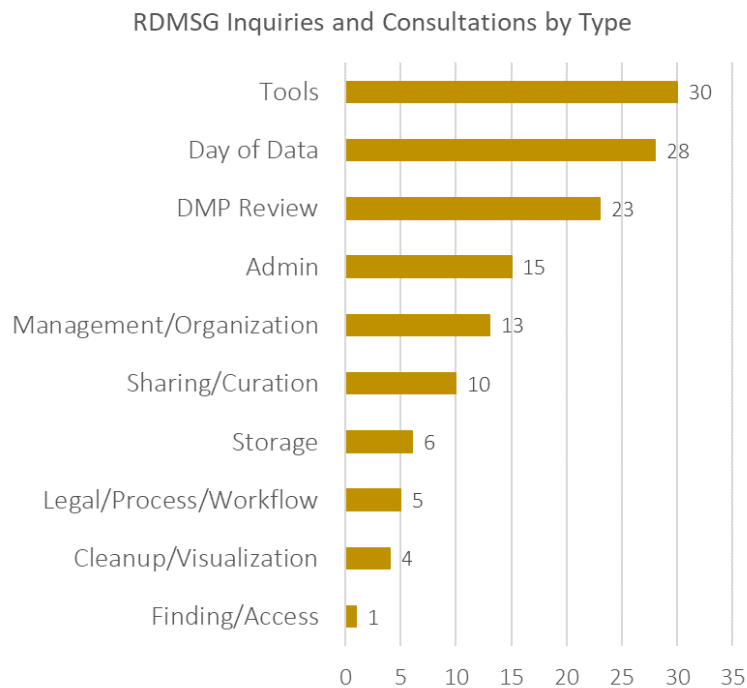| Type	Count
Tools	30
Day of Data	28
DMP Review	23
Admin	15
Management/Organization	13
Sharing/Curation	10
Storage	6
Legal/Process/Workflow	5
Cleanup/Visualization	4
Finding/Access	1

Figure 3: RDMSG Inquiries (including administrative and those related to the 2019 Day of Data Event), and consultations, by type, showing the wide range of data management needs across campus. Not all consultations had "type" recorded so data does not reflect the full set of 2019-2020 consultations.

Consultations by Affiliation

Affiliation	Count
CALS	23
CAS	13
CHE	10
ENG	9
Other	6
Non-Cornell	5
Johnson	5
CUL / CU Press	4
VET	3
ILR	3
LAW	2

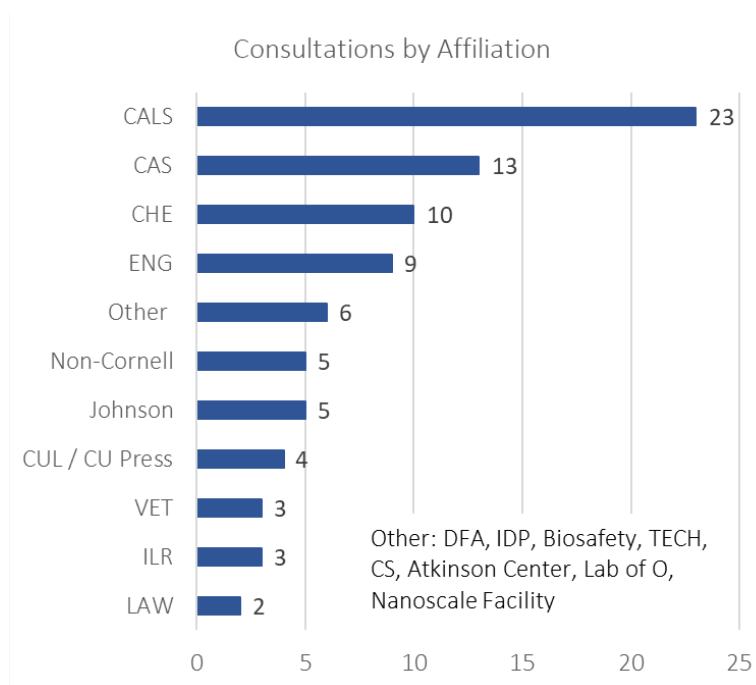Other: DFA, IDP, Biosafety, TECH, CS, Atkinson Center, Lab of O, Nanoscale Facility

Figure 4: RDMSG Consultations by college affiliation. Data reflects only consultations where college was noted, not the full set of 2019-2020 consultations. "Other" category also includes consultations with individuals outside of Cornell University.

<u>Plans for 2020-2021</u>

- Continue consultation efforts in support of agency data management plan preparation and other data-related needs.
- Meet regularly as a group to share consultation experiences, with a focus on expanding understanding of tools, services, policies and practices to build on existing expertise within our current group of consultants.

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 sharing research results. Since then, several other major funding agencies, including the NIH (2003), the NSF (2011) and many others require data management plans with proposals, and stipulate that sharing of research data (with exceptions for confidentiality, privacy, IP and commercial concerns) is a requirement of receiving funding. In addition, many journals (Science, SpringerNature, AGU Publications, PlosOne and others) also require sharing of data as a stipulation of article publishing.
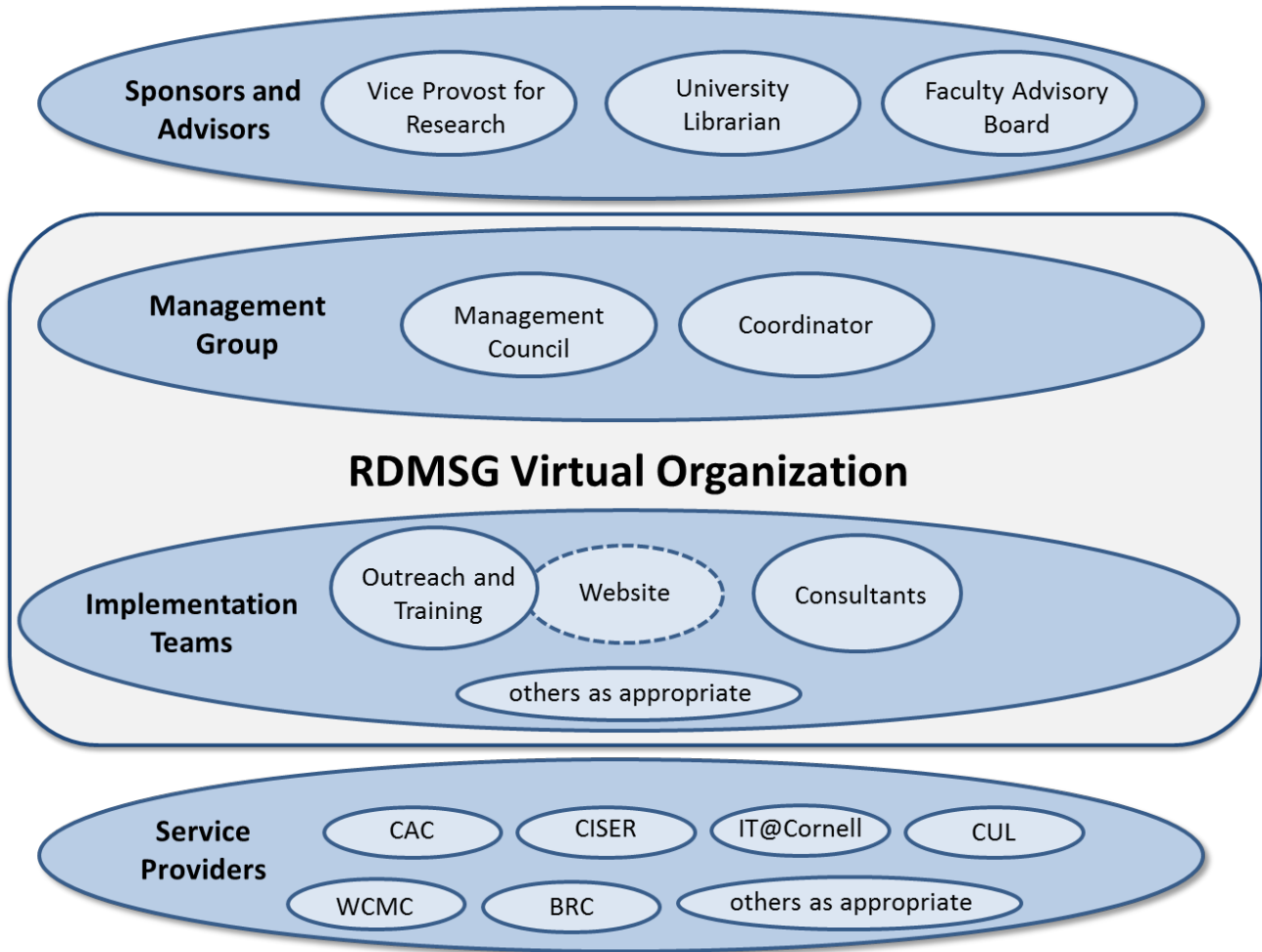
While the RDMSG continues to consult on preparation of data management plans, there is also a clear need to support active data management, and data sharing. These activities are often not "one-time" interactions, but rather an iterative process by which the researchers learns what they need and how to best make that happen. The RDMSG strives to offer multiple ways to facilitate that learning, including comprehensive, up-to-date, self-service information on the website, and accessible and relevant consultative expertise. The RDMSG offers researchers support, but those interactions in turn allow us to gather information about needs and practices of Cornell's researchers, ensuring that campus service providers know what is needed to meet researcher requirements across our campuses.

Finally, continued interest and attendance at workshops and training sessions by staff and students, indicate that they too acknowledge the importance of good and practical data management skills. Such practical data management skills are not necessarily inherent in traditional research pedagogy, and the RDMSG continues to provide supplementary and alternative sources of information by which they can acquire those skills.

> During 2020-2021, the RDMSG will focus efforts on:
> - Increasing basic awareness of the group, and data management activities, practices and trainings through dissemination of a monthly newsletter to the rdmsg-announce email list.
> - Develop and market educational materials appropriate for virtual dissemination, for both synchronous and asynchronous learning.
> - Adding and updating relevant content to the RDMSG website.
> - Organizing and hosting a second Day of Data event.

Appendix A:
RDMSG Organizational Structure

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

RDMSG Virtual Organization

Management Group
- Management Council
- Coordinator

Implementation Teams
- Outreach and Training
- Website
- Consultants
- others as appropriate

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

Appendix B:
RDMSG Faculty Advisory Board Members

John Abowd (ILRLE)
Myles Gideon (ORIA)
William Arms (CS)
Richard Burkhauser (PAM)
Claire Cardie (CS)
James Cordes (ASTRO)
Art DeGaetano (EAS)
Mary-Margaret Klempa (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)
Danianne Mizzy (CUL-IT)
David Lifka (CU-IT CIO, CAC)
Vinay Varughese (WCMC-IT)
David Vernon (CIT)
Wendy Kozlowski (ex officio, RDMSG Coordinator)