TO: David V. Rosowsky, Provost and Senior Vice President

FROM: Mike Austin, Director of System Administration, Enterprise Technology Services
Chris Burns, Library Associate Professor, Chair - Faculty Senate Research, Scholarship and Creative Arts Committee
James Bagrow, Assistant Professor, Department of Mathematics & Statistics
Lesley-Ann Dupigny-Giroux, Professor and Chair, Department of Geography
Cindy Forehand, Dean, Graduate College
Donna O'Malley, Library Associate Professor, Dana Medical Library
Russ Tracy, Professor, Pathology and Laboratory Medicine

DATE: September 30, 2017

SUBJECT: Report of Ad Hoc Data Management Committee

The Ad Hoc Data Management Committee is pleased to share with you our report in response to your charge of March 24, 2017. The report gives a brief overview of data management, addresses the three main tasks outlined in the charge, and makes four recommendations for moving the University forward. In the charge, you assigned the committee the following three tasks:

> *1. Review a recent analysis of data management plans and data management practices on campus.*
>
> *2. Conduct a thorough examination of data management plans in place in select peer and comparator institutions; identify common elements; broadly describe different approaches and their strengths and weaknesses.*
>
> *3. Develop standard (and, where appropriate, agency specific) UVM data management plan language for use by faculty members in the submission of grant applications requiring DMP statements.*

Our recommendations are informed by the first two tasks, an environmental scan of data management plans and data management practices at UVM and peer and aspirant research institutions. There was not enough information available to develop standard data management plan language (in fact the varied requirements of different funding sources and likelihood of on-going changes within each source, make the development of standard language of long-term use problematic); instead we are recommending some practical next steps: to implement a tool (DMPTool, https://dmptool.org/) to support the creation of data management plans, to create a research data support team, to engage more fully with the faculty to identify data management needs, and to work with relevant stakeholders to ensure the University has appropriate data management policies in place.

Overview

Faculty inevitably generate data in the process of conducting their research. These data, almost entirely digital, present significant challenges to ensure that they remain accessible and secure

over their lifespan. While we may be tempted to think of large datasets when talking about research data, smaller datasets present challenges as well and may be receiving even less attention at the institutional or disciplinary level. In addition, data sharing, where possible, is now widely viewed as an important part of scholarly communication.

In 2013, these points were acknowledged and put into a federal directive by the White House Office of Science and Technology. The OSTP directive instructed funding agencies to come up with a strategy for addressing data management concerns. NIH and NSF already had in place mandates for data management plans, and this is the path other federal funding agencies took in response to 2013 OSTP directive.

The Committee reviewed these federal agency requirements for data management plans, with a particular focus on the largest funders of research at UVM (NIH, NSF, DOD, DOE, DOEd, NASA, USDA, DOT). While the agencies vary in how much detail they request and there are agency-specific requirements, our main finding is that the type of information they request generally falls into the following five areas:

1. Types of data produced
2. Standards for data and metadata, where there are standards
3. Policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements.
4. Policies and provisions for re-use and redistribution.
5. Plans for data archiving and preservation and access

Charge 1 - Recent analysis of data management plans and data management practices on campus

Elizabeth Berman, former UVM Libraries faculty member, conducted research on the data management practices of selected UVM NSF grant recipients, analyzing 35 data management plans (DMPs) for the period of 2011-2014. Her analysis of these DMPs showed:

- most DMPs listed multiple types of data (e.g. experimental data, images, samples and specimens, surveys, software, geospatial data, etc.) that the research generates or collects
- 51 % of DMPs addressed metadata and data documentation
- 43% of DMPs addressed storage; 40% of DMPs addressed storage and back-up protocols
- 80% of DMPs indicated willingness to share data
- 100% of DMPs addressed data preservation;
- 49% of DMPs specifically mentioned subject data repositories

These results show mixed compliance with NSF requirements. While not addressing all of the requirements in the data management plans did not impact the ability of these researchers to obtain funding, it is a good indicator that best practices for research data management are likely not being followed once the research has begun, possibly putting that data at risk. Berman noted the following two points as key obstacles to this compliance:

1. "The role of **education and outreach** will be critical in any effort to support data management on campus. Not all faculty understand the need to preserve or share research data, and so there is a need to raise awareness to the issues of Open Data and data management planning."

2. There is "…general confusion about **UVM institutional support for research**. Because of the decentralized nature of the institution, there is no single place for a researcher to find information regarding data management planning; information that can be found is often so out of date that it is no longer useful, while other relevant information is well hidden. Faculty don't always know who to talk with to get answers; as a result, there are many misconceptions and misunderstandings about the available infrastructure and services for research faculty."

As Professor Berman noted, there is some support at UVM for research data management and the creation of data management plans, but it is not well documented or well-coordinated among administrative units. Some examples of support are listed below:

- Libraries Guide
  http://researchguides.uvm.edu/datamanagement
- Sponsored Programs Administration
  https://www.uvm.edu/spa/?Page=dataacquisition.html
- IRB Guidance on Data Management in Human Subjects Research (Appendix O, p. 128)
  http://www.uvm.edu/irb/Research%20and%20Policies%20and%20Procedures%20Manuals/researchmanual.pdf
- College of Medicine Biostatistics Core Facility
  http://www.med.uvm.edu/uvmcancercenter/core-facilities/biostatistics
- ETS/Vermont Advanced Computing Core
  http://www.uvm.edu/~vacc/
- ETS Research Storage
  https://www.uvm.edu/it/about/services/?Page=storage.php

Educating faculty about research data management, including better communication about existing institutional support in this area, could go a long way to improving both the quality of data management plans and data management itself at the University.

Charge 2 – Conduct a thorough examination of data management plans in place in select peer and comparator institutions; identify common elements; broadly describe different approaches and their strengths and weaknesses.

Due to the lack of availability of individual data management plans at peer and comparator institutions, we focused our assessment on the support for research data management, and more specifically data management planning, at these institutions. Reviewing institutional websites, we looked at data management plan support, data policies, DMPTool status, data storage, local repositories for preservation and sharing, services provided at the unit level, and the availability of consultations and workshops (see attached spreadsheet). From our study of ten peer and eleven aspirant institutions, gathered from a list of research institutions provided by the OVPR

(see attached list), we can draw some broad conclusions and point to some potential models for UVM.

Conclusions

1. Support for research data management and data management planning varies by institution, but it is almost always handled by information technology, the library, the office of research/sponsored programs, or some combination of the three. Libraries are the most consistent source of information, almost all that we reviewed provide a guide to data management planning, which generally does the following:
   - discusses the common elements in any data management plan
   - links to agency requirements
   - links to the DMPTool (a free tool to assist in creating data management plans, which offers an opportunity for institutions to become members and create customized templates for their researchers).
   - lists resources at the institution and elsewhere for data sharing, storage, and preservation.

2. Almost all institutions have policies relating to research data management. Guidance is found in policies that are explicitly about research data management as well as about narrower and/or related topics. These policies touch on data use, ownership (intellectual property), retention, access, security, and IRB issues. UVM does have policies and guidelines in place that touch on research data, such as the Intellectual Property Policy and the Manual for Human Subjects Research, but appears to be less explicit policy-wise than some of our peer and aspirant institutions, although the current Intellectual Property Policy indicates that a Research Data Retention Policy is under development.

3. The DMPTool has become the *de-facto* tool in the United States for supporting the creation of data management plans. Only one of the 21 institutions we looked at did not point to the tool. Many institutions take the additional steps of becoming "partners" (DMPTool Institutional Members) and customizing the tool. 15 (10 aspirant, 5 peer) of our 21 peer/aspirant institutions are partner institutions and 7 (5 aspirant, 2 peer) of those provide the additional service of customized templates with institution-specific information. This 90-second video is an excellent introduction to the tool - https://dmptool.org/video.

Models

We found strengths and weaknesses at all of the institutions we looked at. A few appear to share one of our main weaknesses: not clearly communicating information about institutional support for data management. However, many institutions have made progress on this issue, as well as other data management issues, and serve as good models for UVM. Below, we highlight some ways in which three of these institutions are supporting data management.

Virginia Commonwealth University

- Detailed policy on Research Data Ownership, Retention, Access, and Security. https://policy.vcu.edu/sites/default/files/Research%20Data%20Ownership,%20Retention,%20Access%20and%20Securty.pdf

- DMPTool partner with customized templates (pointed to from Library and Office of Research and Innovation). Example of a DMP created with a VCU template - https://dmptool.org/plans/15751.pdf
- Use of Open Science Framework to support research and an institutional repository, Scholars Compass, to host and share datasets.
  https://osf.io/institutions/vcu/
  http://scholarscompass.vcu.edu/
- Library Guide with general and institution-specific resources.
  http://guides.library.vcu.edu/data

University of Missouri, Columbia

- Good information about research computing on campus, with clear information about general purpose research storage, including pricing, and a page linking out to other campus resources, such as the Library. https://doit.missouri.edu/services/research/
- Library guide with good general information about data management and data management plans, which links to other campus resources and which their Sponsored Programs office links back to.
  http://libraryguides.missouri.edu/c.php?g=28117&p=173327
- Library guide contains information about general and discipline-specific data repositories, and also points to their institutional repository, and includes language about their repository which can be plugged into a data management plan.
  http://libraryguides.missouri.edu/c.php?g=28117&p=173329

Tufts University

- Research Data Management System provides researchers with tools to address unified data management, data curation, and compliance with federal grant data management mandates. https://it.tufts.edu/rdms
- Data management guide and consulting services from Library - https://tischlibrary.tufts.edu/services/data-management
- DMPTool partner.
- Detailed research data storage information. https://it.tufts.edu/r-drive

Charge 3 - Data Management Plan Language

The rapidly changing landscape of data management requirements makes the development of standard language for long-term use unsatisfactory. Rather, we need to establish a flexible approach that will lend itself to such a mutable environment. Our review of peer and aspirant institutions demonstrated that most institutions either simply point researchers to DMPTool or have taken the additional step of becoming DMPTool partners, thereby adding the ability to provide customized institution-specific templates. The committee believes an easy way to move UVM forward is to become a DMPTool partner and create customized templates within the tool. There is no cost to become a partner. An important step after that would be to gain feedback from faculty on what types of information to include in the templates. This might include suggestions about data storage, sharing, and preservation, both at UVM and beyond, and could be modeled on the many DMP examples that have been made public through DMPTool, and sites like those listed below:

- ICPSR
  https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/resources.html
- DATAONE example plans
  https://www.dataone.org/data-management-planning
- NSF Engineering Data Management Template, University of Michigan
  http://hdl.handle.net/2027.42/86586

Recommendations

1. Become a partner institution of the DMPTool. There is no fee to become a partner institution. The advantage of using this tool is that it provides up-to-date agency-specific templates for creating data management plans, which can be customized with UVM specific information if we are a partner institution, making the process of creating these plans easier for researchers. Having access to assistance and the most accurate institutional information would also lead to more accurate and higher quality management plans. Adoption should be followed by promotion and training to ensure that faculty are aware of the tool and comfortable using it. The tool should be periodically assessed to ensure that it meets the needs of the University and its researchers.

2. Create a research data support team led by representatives from Enterprise Technology Services, the Libraries, and Sponsored Programs, with faculty advisors from across the University. Cornell's Research Data Management Services Group is one possible model - https://data.research.cornell.edu/content/about.
   Charge this team with coordinating information about data management plans for researchers at UVM, centralizing information found on the Libraries guide and elsewhere on the UVM website. Assign ownership for maintaining this information and promote and update it on a regular basis. The outcome should be that researchers find it easy to identify and get support for data management, including the creation of data management plans.

3. Assess further faculty data management needs in relation to data management plans and data storage, preservation, and sharing. The assessment should expand on the work done by Elizabeth Berman, which was intentionally limited in scope, as well as surveys at other institutions, such as one recently conducted at the University of Central Florida - http://www.ist.ucf.edu/hpc/rcd/Beile_datahandout.pdf.
   The assessment should gather information from researchers across campus, who receive funding from a variety of agencies, and provide a broader and more current sense of data management practices at the University. The assessment should be designed to inform data management plan templates, guidance provided by administrative units, and as evidence for any new proposals that might involve investments in financial or human resources for tools and services to support data management at the University.

4. Review current UVM policies related to research data and identify and fill any policy gaps. The current Intellectual Property Policy refers to a Research Data Retention Policy that is under development, so part of this may be underway already. This document from Wayne State points to the variety of policies/guidelines that touch on data management at that institution - http://rds.wayne.edu/pdf/WSU_Research_Data_Policies_v5.pdf.