Library Faculty Fellow Report: Assessment of DUL and its role in the Research Data Management Practices and Needs of Drexel Researchers

Christopher Sales
Spring/Summer 2018

Introduction

This Library Faculty Report describes my assessment of Drexel University Libraries (DUL) efforts to improve their support services for research data management (RDM). My goals were to aid in the interpretation of information collected by DUL on the current state of RDM practices and needs of faculty across Drexel, collect information on the RDM practices and needs of research faculty in the School of Biomedical Engineering, Science and Health Systems, and to make suggestions on how DUL could raise awareness to faculty and students the RDM support services that they provide.

Faculty Survey on Research Data Management

The Report of 2017 Faculty Feedback on Research Data Management (RDM) by Dean Danuta Nitecki and librarian Deb Morley (from here forth referred to as the "Report") provided a summary on responses to a survey conducted by the Data Stewardship Forum at Drexel University to assess the research data management practices and needs of Drexel Faculty. Two major takeaways from the Report and my re-assessment of the survey responses, especially the free responses, were (1) the needs to improve education on agency requirements and workflow procedures for research data management, and (2) the low number of faculty that use archival repositories, 11% (n=13) of respondents according to Question 13 and 25% (n=29) according to Question 14. The first takeaway is striking because the primary respondents to the survey were faculty, who are cognizant that they are responsible for managing research data for their research groups (Question 18) but who acknowledged that they did not know what RDM requirements were mandated by their funding agency (Question 14). Some of the conclusions of the Report relate to this need to improve education among faculty (and their students and research staff) regarding RDMs, namely the need to "increase awareness among researchers of the requirements for RDM, the potential liabilities the University [and faculty] face with inadequate RDM practices, and the support services available to enable their productivity" (p. 5).

These findings from the Report and my assessment of the free responses of the survey contribute to some of my suggestions in my constructive critique of the Drexel University Libraries (DUL) webpages on RDM.

Data Management Plans and Interviews Regarding RDM practices by Biomedical Faculty

After attempts at contacting the Associate Dean of Research in the School of Biomedical Engineering, Science and Health Systems, Kenneth Barbee, and e-mailing research active faculty in Biomed (tenured, tenure-track and research professors, including affiliated faculty), I was only able to receive two examples of data management plans (DMPs) and was only able to conduct one in-person interview. The DMPs I received were from Gail Rosen, who is an affiliated faculty in Biomed and an Associate Professor in the Department of Electrical and Computer Engineering, and Gregory Fridman, who is an Assistant Research Professor in Biomed.

Gail Rosen provided a DMP that she submitted to NSF in 2016 for National Research Traineeship (NRT) grant. This DMP (see Appendix A.1) was extremely thorough delineating between the types of data that would be generated and used (research vs. assessment data), as well as containing clear sections for Research Data, Assessment Data, Format/Content Standards, Access/Sharing, Reuse/Redistribution, Data Storage and Preservation, and Investigator Roles. This DMP includes information on archival repositories that will be used, as well as computing infrastructure (namely the Proteus computing cluster at Drexel) that is available for use for the project associated with the DMP. It also describes how source codes used for analyses will be shared in an open-source manner.

Gregory Fridman shared his DMP that was submitted to the Partnerships for Innovation-Research Partnerships (PFI-RP) Program at NSF. It is also fairly detailed, describing the roles of the investigators and collaborators (including Danuta Nitecki), as well as detailed sections on Expected Data to be Managed, Data Formats, Period of Retention, Data Storage and Preservation, and Sharing and Access of Primary Data. One interesting thing that this DMP points out explicitly is specific types of data, materials, and collections that will not be managed according to this DMP (e.g., preliminary data, raw data, drafts of science papers, trade secrets, physical samples). This is interesting because, from my best understanding, most faculty do not typically include these phrases in their DMP.

In addition to providing me with a copy of his DMP, Gregory Fridman also allowed me to meet with him to ask questions about his RDM practices and needs. Dr. Fridman is the Co-Director of the Plasma Medicine Laboratory in the Nyheim Plasma Institute of Drexel University, located in Camden, NJ. Similar to the RDM survey responses from 2017, Greg said that he uses cloud computing frequently to store his research data but that the sharing platform that he uses is dependent on who he is sharing the data with, as he uses Dropbox for sharing data with external collaborators but is now using Drexel's access to OneDrive to share with data with Drexel students working in his lab. In addition to using cloud computing, he also shares large amounts of data (>100 TB) using physical hard drives because it is faster to mail the hard drive

than wait for the data transfer over the internet. Like many of the survey responders, Dr. Fridman states that the RDM needs of each project are unique and that it is his responsibility as the PI to determine the best DMP for each project. For archiving data, Dr. Fridman uses network attached storage devices. Some of the storage devices are password protected and access to them physically is secured with a lock, due to requirements of a particular federal grant. In relation to how he believes Drexel can best support his RDM needs, Dr. Fridman wants the University to supply resources for cloud integration for storage, archiving, and sharing data but he is worried that the university will try to enforce policies relating to the use of these resources that might interfere with research productivity. Another area that Dr. Fridman believes the University could do better in relation to RDM is educating faculty on what constitutes best practices in RDM.

Assessment of DUL RDM Webpages

The DUL website contains a set of webpages relating to RDM. Navigation to these webpages from the main landing page for DUL can be made by hoovering over "Services" at the top of the page to access its dropdown menu. Although it is logical that webpages related to RDM support would be located under Services, I initially did not look for it there. Instead, I used the website's Search feature to find the RDM webpages.

From the main landing page of DUL, when you click on "Services" at the top of the page, it brings you to the "Overview" page for Services. On this Overview page, in addition to a link on the menu on the left hand side of page, which contains a link to "Research Data Management Support," there is also a link in the main text of the page "Data Management services for faculty & researchers" that brings you to the same RDM webpage. One shortcoming that I found while navigating through the Services Overview page is that if you click on the link for "Faculty" in the main text, the page it brings you to does not contain in its main text information about RDM services offered by DUL (even though it is linked on the left-hand side of the page. In addition, on this Faculty page, on the right side where it says "Faculty Links," it would be good to include a link to RDM support services there too.

The Research Data Management Support page has five links in the main text:

1. Guidance on data management plans required for funded research
2. Support in using DMPTool
3. Assistance in acquiring persistent identifiers for publications (such as DOI)
4. Consultation on how to share data to meet publication requirements
5. Training on research data management.

These five links correspond to the same links that appear under Research Data Management Support in the menu on the left-hand side of the page:

1. Data Management Plans
2. DMPTool
3. DOI Services
4. Sharing Data
5. Training

When you navigate to the Data Management Plans page, the page that comes up describes that there are RDM requirements for some funding agencies and provides a link to DMPTool. This page is extremely bare bones. It would be beneficial for it to include links that describe the RDM requirements for different funding agencies, as well as give examples of DMPs for specific funding agencies. For instance, it could link out to the Funder Requirements page on the DMPTool site that contains a table with templates of DMPs and links to funding agencies guidelines for DMPs. DUL could also create a table similar to NYU Libraries that describes in simple terms if any agency requires a DMP, while also linking out to each agencies guidelines on DMPs. In addition, DUL could link out to Sherpa Juliet, which allows users to search for funders policies on open access, publication and data archiving

The Data Management Plans page on DUL's website should also contain links to sample DMPs, or at least mention that public examples of DMPs are available on DMPTool. NYU Libraries webpage on Data Management Plan contains a slideshow "Why a Data Management Plan?", which contains bullet points that gives reasons researchers should use a DMP – a similar, eye-catching graphic on DUL's website would be useful. Another useful thing to include on this page would be something similar to the University of Pennsylvania's library webpage that includes a breakdown of "Best Practices in Data Management," on topics such as "Writing a High Quality Data Management Plan," "File Organization," "Documenting Your Data," and so on.

The second main link on DUL's RDM support webpage is for DMPTool. Although I had heard about DMPTool from the College of Engineering's Library Liaison, Jay Bhatt, I have not taken full advantage of the features of DMPTool. One useful feature that is not being fully utilized at Drexel is the ability to publicly share DMPs with other Drexel Researchers. Currently, there are five shared DMPs at Drexel but only one actually contains a completed DMP (Shannon Capps DMP for NASA). If we can encourage more faculty to share their DMPs on DMPTool, it would help faculty new to DMPs create them for their grant proposals. While looking through the features of DMPTool, one question I had was, Does DUL have statistics on the number of its users from Drexel?

The third main link on DUL's RDM support webpage is for DOI Services. This webpage is useful because it shows that DUL subscribes to services that helps Drexel researcher obtain DOIs.

Since I don't see myself using the DOI Services offered, I am curious as to whom at Drexel has used this service. To draw more users to this service, it might be interesting to have a ticker or short list on the page of the most recent DOIs created by Drexel researchers (that are not associated with peer-reviewed journal articles or books). Another suggestion for this page is to include a link to Drexel's institutional repository and archive, iDEA.

The fourth link on DUL's RDM support webpage is related to Data Sharing. This webpage is fairly bare. It contains an extremely useful link on to the DUL guide on "Research Data Repositories." This guide on Research Data Repositories has a lot of useful information and links that should be moved to the DUL page on Data Sharing. For example, some of the information on Dryad, Figshare, and Harvard Dataverse would be good to have on the Data Sharing page. The Data Sharing webpage should also include information on infrastructure available at Drexel for researchers to share their data internally with other Drexel faculty and students, as well as means for sharing securely with external collaborators (e.g., Figshare). Another thing missing from this webpage is information on RedCap. The Drexel University College of Medicine (DUCOM) has an informative webpage on accessing REDCap for their faculty, students, and staff.

The fifth link on DUL's RDM support webpage is for Training. The term "training" is vague and it's not intuitive what information this page would contain. The page does contain links to Stanford University Libraries guide to Data Best Practices, as well as primers on RDM provided by DataOne and ICPSR (the Iner-university Consortium for Political and Social Research) I think it may be better to call it "Data Management Best Practices" and for it to include workflow diagrams related to RDM best practices.

Lastly, one simple suggestion, for all of the DUL RDM support webpages, is to include names and faces of staff at the library that respond to the email datamangement@drexel.edu, in order to make the services provided seem more personable. Since faculty are the primary persons responsible for managing their research data and the fact that the management of research data is essential to their research group's productivity, knowing who they can reach out to personally (especially when they are in the middle of writing proposals) for help with their RDM needs would be comforting and reassuring.

Concluding Remarks

Although only 192 faculty (13.4% of the total faculty) responded to the 2017 survey on RDM, it is evident that there is a need to better educate faculty and their students of agency requirements, best practices, and support services available to them related to RDM. DUL can play a major role in further educating faculty and students about RDM. Most faculty are unaware when writing proposals that DUL offers RDM support services. On Drexel's Office of

Research website, there is no information on how to prepare DMPs or how to find RDM guidelines for each funding agency. Nor does the Office of Research website link out or mention that DUL provides support services on RDM. It is a great benefit to Drexel researchers that DUL has created a series of webpages on RDM but these pages could be greatly improved. Not only could the appearance and content be improved but DUL should also work on raising the visibility to Drexel researchers that they provide RDM support services. Although faculty are usually the primary party responsible for RDM, it is often their students in their research groups that generate, handle, and process the data. Therefore, it might be worthwhile for DUL to work with Drexel's Graduate College on developing classes, workshops, or a course on RDM that complements their efforts on teaching graduate students about reproducible research. If DUL takes an active role in development of teaching material for graduate students to learn about RDM, it could also naturally lead to improvements to DUL's webpages on RDM that I have noted above.