Research Data Support at University of Wisconsin-Madison: A Case Study

Cameron Cook, Jennifer Patiño
University of Wisconsin-Madison
cameron.cook@wisc.edu, jennifer.patino@wisc.edu

Abstract:
This brief case study will introduce the research data sharing and compliance landscape in the United States (US) and the ways in which the University of Wisconsin-Madison Libraries are responding to it through their local research data support services. This study provides a brief introduction to the federal policy and data sharing context in the US, describes structure and services provided by the UW-Madison Libraries' Research Data Services, discusses the ways the libraries are facilitating compliance with data sharing requirements through supported research data repositories, and concludes with recommendations based on lessons learned from the development and ongoing support of these services.

Introduction

Research data support at the University of Wisconsin-Madison (UW-Madison) has been greatly shaped by the federal research data policy in the United States (US). As a large institution with a high amount of federal funding, many of our researchers are subject to policies that require data management plans and public access to research data at time of publication or the close of their projects. These requirements are continuing to evolve, new policies are emerging, and the culture around data sharing continues to shift in higher education. Given this complex and frequently changing environment, the UW-Madison Libraries have aimed to be responsive to the pressures researchers are under and create an ecosystem of services that can support the campus in being compliant with funding agency policy as well as promoting the use of best practices throughout the research data lifecycle.

Federal Research Data Policy in the United States

The current state of UW-Madison's research data support and data repository options has been shaped by a memorandum released in 2013 from the White House's Office of Science and Technology Policy (OSTP) and its subsequent impact on the campus community. The memorandum required that the publications and underlying research data funded by a specific subset of federal funding agencies must be publicly accessible, meaning available to the public in digital formats and online for free, within a year of publication and each agency was directed to develop individual policy in support of this aim[1]. When those agencies created and implemented

[1] "Increasing Access to the Results of Federally Funded Scientific Research," President Obama White House Archives, February 22, 2013, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

their policies in response to this memorandum, many of them began to require data management plans as part of the proposal process. These data management plans were to detail how data would be appropriately managed and organized as well as where data would be made publicly available. As a very large research-intensive institution federal grants make up a significant portion of the yearly budgets and these changes directly affected many of the campus' researchers. This created a new and pressing impetus for managing and sharing digital data on the campus which Research Data Services, discussed in more depth later, responded to by offering free guidance, training, and consultations to support these new needs.

In recent years, publishers in the United States have also begun to support the global cultural shift towards sharing research data. Many journals now require that research data underlying accepted publications must be made publicly available prior to the publication of the article. In turn, there has been a noticeable anecdotal increase in requests from campus researchers to help share their data specifically for publication. Between the federal agency policies and publisher requirements, there has been recent increased buy-in and interest in Research Data Services from our campus partners.

Researchers on the UW-Madison campus will undergo another wave of change in data sharing expectations in 2023 and beyond, as new policy from a funding agency comes into effect and the details resulting from a new White House OSTP memorandum are revealed. In January of 2023, a new data management and sharing policy is going into effect from the National Institutes of Health (NIH). This new NIH policy now applies to all NIH researchers with scientific data, as opposed to the former iteration that applied only to researchers with large awards. It is more stringent and affects a large portion of campus researchers across many schools and colleges. This significant policy change is drawing renewed attention for data management support and has increased demand for training and outreach from Research Data Services in preparation for the shift. Alongside this, the White House's Office of Science and Technology Policy released a new memorandum in August of 2022 that will change expectations for public access to research publications and data moving forward. This new memo expands the requirements to all federal funding agencies and requires that publications and data be immediately available upon publication without embargo[2]. This is a big change for researchers, agencies, libraries, and publishers in the United States. The memo details a number of other requests for agencies to consider in their policies as well as charges a subcommittee to work on a number of related items to facilitate these new requirements. Until the affected agencies release their specific policies, it is hard to anticipate how to adjust research data support services appropriately. However, Research Data Services and the campus libraries are staying abreast of updates, conducting outreach across campus to raise awareness, and are prepared for this to shape strategies and near

[2]"Ensuring Free, Immediate, and Equitable Access to Federally Funded Research," White House, August 25, 2022, https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

term goals for the next few years. This entire policy landscape positions UW-Madison libraries research data support in an interesting confluence of a rapid shift toward open access and open data with the challenging realities of how to support a large, decentralized campus as these policies go into effect. Research Data Services have been a low-cost way to support researchers locally through focusing on developing cross-campus relationships and relying on collaborative expertise to provide lightweight services. The Libraries and Research Data Services are also engaged in identifying an array of data repository options that will help researchers across a spectrum of data needs to share their data appropriately now and into the future.

Research Data Services - An Overview

Research Data Services (RDS) is a completely free resource to anyone in the UW-Madison community. It is open to all faculty, staff, and students that support data management and sharing needs. The service aims to help "researchers as they gather, manage, and share their research data to make their data citable, reproducible, and publicly accessible."[3] RDS advertises three main services to campus researchers: data management plan review, consultations, and education and training. However, the RDS team also spends a large amount of effort behind the scenes to provide informational resources and best practices, conduct outreach to the campus community, maintain and grow a network of campus partners, and advocate for known needs or gaps in research data support on campus.

RDS is a collaborative effort, led by the Libraries, and the team is an interdisciplinary group of experts from across the UW-Madison campus. This interdisciplinary focus allows RDS to leverage the expertise of colleagues and partners to help researchers across the life cycle of their projects. Members of the team include librarians with expertise in data management, outreach, instruction, and subject area expertise in data curation, the humanities, science and engineering, the health sciences, social sciences, and geospatial data. RDS also has a fundamental partnership with the Research Cyberinfrastructure initiative within the campus' Division of Information Technology (DoIT). The Research Cyberinfrastructure team has expertise in research information technology and infrastructure, research computing—including cloud computing, and solutions for storage, backup, and internal data sharing. This team provides critical expertise, supports the RDS website, and partners with RDS on special projects. RDS also has relationships with the University's Data Science Hub, which provides support for training in computer science and data science. They provide expertise in data science facilitation, instruction, project support, and help with machine learning.

As mentioned, RDS provides consultations, both one-on-one with researchers and in group settings with labs. These consultations seek to assess data management needs and offer

[3] "Vision & History," Research Data Services, n.d., https://researchdata.wisc.edu/our-services-2/vision-history/

recommendations while also guiding researchers to useful campus resources where appropriate. Developing a robust network of campus partners has been extremely beneficial to being able to make referrals to campus resources for a wide array of needs during these consultations.

The education and training that RDS provides takes a number of different forms. Currently, a large amount of effort is going into education for the upcoming NIH policy changes and is trying to reach beyond researchers to research administrators, assistant deans for research, departmental IT services, the institutional review boards, and more. RDS can provide training for labs and for classrooms in best practices for data management, data sharing, and reproducibility. It also offers on demand, self-guided training modules on data management basics on the RDS website.

Finally, RDS offers assistance with data management plan review. Creating data management plans is a good practice for researchers in general, but as mentioned, many funders require researchers to write detailed data management plans to assure that their data will be protected and properly cared for at all stages of the research project. RDS offers to review those plans before they are submitted to the funding agency with the grant application to help ensure that the plan meets the agency's requirements, any legal and university requirements, and generally uses good data management practices. This has been especially helpful for researchers as it is often difficult for them to keep up with the changing funding agency requirements and to know which resources at the university will help them meet those requirements.

In terms of reviewing data management plans, RDS leverages a free tool called DMPTool (https://dmptool.org/plans). This tool is extremely similar to DMPonline, but has a United States funding agency focus. It has templates available for many of the most common funding agencies, which makes it easier for researchers to write plans that will meet requirements they need for that specific grant. It also has the option for researchers to create a data management plan with a general template. Once researchers write the draft of their plan, they can request feedback directly within DMPTool. A notification goes to the team that there is a new plan to review and a consultant from RDS is able to review the draft and provide feedback on the plan through the tool. DMPTool enables the data management plan review service to be quick, efficient, and easy for researchers.

As noted earlier, behind the services advertised to researchers, RDS takes on a number of other activities to enable robust research data support. The RDS group meets regularly to share knowledge, build community, and keep each other informed of updates in each consultant's area. Building and maintaining community is key to having a referral network that can be leveraged to direct researchers' questions to the person who will be able to provide them with the best support. This also spreads the burden of knowledge among the group, relying on everyone's individual expertise across the lifecycle to make up a collective knowledge that is more robust.

RDS also collaborates on outreach—an important part of the services' efforts. It is critical to get information about RDS services, resources, and updates in funding agency requirements to researchers at the point of need so that researchers can be proactive with data management and sharing practices from the beginning and have what they need to be compliant with their requirements. Outreach to students is also important so that they know about data management practices early in their careers and can develop skills they need to be successful in their research projects later on. Outreach methods include tabling at student events, presenting at campus events and conferences, creating fliers about services, maintaining a special topics and news blog on the website, a regular email digest with information on workshops and news items, social media presence on Twitter, and organizing workshops and events.

As mentioned, RDS relies deeply on collaborative partnerships to provide support. Through such collaborations, RDS is able to take on special projects that fill known needs on campus. A few examples of these include the Data Storage Finder (https://storage.researchdata.wisc.edu/), the Researcher Toolkit (https://researchertoolkit.wisc.edu/), and a membership with Dryad, discussed later. The Data Storage Finder is a tool that RDS created with the Research Cyberinfrastructure team. It is modeled after a finder tool by Cornell University and leverages the Drupal module provided by the same institution. The UW-Madison campus has a large number of data storage services available to researchers and researchers often found it hard to navigate information related to those services. This tool aims to improve this and make it easier for researchers to parse through the storage options available to them and find the best solution for them. The tool asks researchers to answer a series of questions about their project needs, their research data, and any regulatory or compliance needs. Based on their answers, the tool recommends data storage solutions available on campus that may fit their needs. Another example is the Researcher Toolkit which was built in collaboration with the Data Science Hub, the Wisconsin Institute for Discovery, and Research Cyberinfrastructure. The toolkit is a guide for UW-Madison faculty, staff, and student researchers that points to helpful resources for each phase of their research project. This tool also aims to make finding research resources easier for researchers by collecting it all in one place and presenting it in a way that makes it easier for researchers to navigate.

Data Repositories at UW-Madison

While RDS always recommends that researchers use appropriate disciplinary repositories first, the Libraries recognize that not all disciplines have repositories and many repositories have limits that may prohibit a researcher from using them. Given this, the Libraries are seeking to create an ecosystem of solutions for data sharing that can supplement disciplinary options and provide support for an array of needs.

The first repository provided to the UW-Madison campus was the open access institutional repository, MINDS@UW. This repository predated the aforementioned policies and was built around 2008. MINDS@UW uses DSpace, which is an open source software and a ready-to-go software for repositories. While there are companies who do provide hosting and support for DSpace, the UW-Madison libraries host it locally. This repository allows all scholarly outputs to be deposited including research data, publications, pre-prints, conference proceedings, posters, and other objects. The structure of DSpace allows researchers to have a community - essentially a landing page - for their department or their lab, and then allows for multiple collections of individual objects within that community.

A number of UW-Madison researchers are producing very large quantities of research data but due to this are limited in their data repository options as they exceed most repositories' data deposit limits. Given this local need, the libraries are partnering with the Research Cyberinfrastructure team to conduct a pilot project. The libraries leverage their data storage infrastructure and a tool called Globus (https://www.globus.org/). The item record and the metadata remain discoverable through MINDS@UW, but the record links out to the data via Globus where it is then accessible for download and use. Currently, the pilot is offering researchers to deposit up to 1 terabyte of data. Progress on this work has been a little slower than anticipated due to the COVID-19 pandemic and limited resources, but its goal is to conclude within two years and to see if supporting this need is useful for campus researchers as well as if it is feasible as a long-term service. MINDS@UW is also currently undergoing some strategic evaluation and some road-mapping is being done for the next few years of the repository to ensure it remains a valuable solution for the campus. The evaluation and roadmap seeks to identify areas of immediate need for viability long term, areas of growth for features, and to identify opportunities for further improvements that can be completed with the Research Cyberinfrastructure team during the pilot project.

RDS and the libraries wanted to provide another, more robust option for a large portion of the campus that has scientific data. The data repository, Dryad (https://datadryad.org/stash), is a third-party open access generalist repository that RDS has been recommending to researchers for a long time. Dryad recently released a new membership model that allowed for an affordable way for the libraries to scale data sharing support more sustainably and provide more robust infrastructure. Through the membership, campus researchers have unlimited deposits of up to 300 gigabytes per deposit. Dryad provides a number of valuable features including, robust storage and preservation infrastructure, minimum curation of deposited data sets, integrations with certain publishers, and the ability for deposited software and scripts to get automatically pushed to another repository called Zenodo, that specializes in that work. Researchers are able to make use of this on-demand without contacting RDS, though RDS is able to provide support and training for it. This makes it very easy and fast for researchers to deposit their data to comply

with requirements while campus can be reassured that data is being deposited into a libraries-vetted repository.

While Dryad and MINDS@UW provide data sharing options for many at our campus there are still gaps in data sharing support. Due to this, RDS provides consultations and education to help researchers identify other repositories that suit their data needs. One example that RDS is conducting significant outreach on currently, is a repository called ICPSR (https://www.icpsr.umich.edu/web/pages/). This repository will be a key resource for local NIH-funded researchers that have social science data or social science methodologies, especially those with human subjects data or other sensitive data as the repository provides different restriction options. Given its potential future value for the campus community, RDS is dedicating significant effort to plant the seed with researchers now that they will need to consider data sharing costs early on as they write their data management plans and request curation cost estimates from ICPSR in advance to be included in the budgets of their funding proposals.

Recommendations

There have been many lessons learned as RDS and data sharing at UW-Madison has matured. For those institutions that are just starting with research data support, there are a few recommendations to share based on local experiences though it is encouraged to use what is useful for the reader's applications.

Research Data Services

First, data management needs can vary widely across disciplines and even throughout the stages of a project. One data librarian can't possibly know everything or be an expert in all aspects. Due to this, RDS strongly encourages connecting with others on campus who have different areas of expertise. In this same vein, a collaborative approach has given UW-Madison the best results. In building a referral network with colleagues across campus, RDS has been able to become a useful first point of contact for researchers looking for support.

Keeping track of researcher needs that surfaced during consultations and discussing them at the regular RDS meetings has allowed RDS to see where it would be valuable to expand services and where there might be significant gaps that campus could help fill.

RDS has also developed strategies for keeping costs low while having high impact. Where possible, RDS relies on free or nominal cost services and tools. By leveraging partnerships with other campus units, RDS has also been able to share costs for support and infrastructure for services that are developed. This is more sustainable long term for all partners involved. Collaborating on projects with other campus units has helped RDS make sure that the time and effort invested in creating resources will be of the greatest benefit to the largest number of people on campus.

Data sharing support

A first and critical recommendation for those starting with data sharing support is to be responsive to the pressures and environment of your researchers and campus. Local UW-Madison solutions have been reactive—having to adjust services or pilot new options based on the needs as they have been encountered. In line with this, it is recommended to use the repository model that makes the most sense for each institution's needs. UW-Madison has used a couple of different options to help meet capacity and different use cases.

Similar to the research data services recommendation, working with other campus groups to address gaps for data sharing on campus is beneficial. Unless an institution has a lot of funding, people, and resources available for a data repository, partnering with other campus services is a great way to share infrastructure and share staff time for your services. It also enables each partner to do what they do best, rather than one unit trying to solve everything on their own. It is also recommended to plan for the growth of the repository at the outset of it. If an institution were to get a lot of demand from the campus, would the storage or service model scale well?

Finally, one area that UW-Madison would like to be able to put more resources and effort into, is robust outreach to our campus researchers. The local repositories could have more impact on supporting data management plans and research proposals, as well as more strategically collect research data if the service was able to conduct more consistent and targeted outreach. Unfortunately, good outreach requires a lot of sustained effort and staff time which is difficult to do with limited resources. If possible, it is recommended that outreach be a strategic priority and that it is considered for outreach to be included in the job responsibilities for a position.

Conclusion

Providing research data support is a challenging and fast-paced effort. However, there are many ways to provide cost-effective and helpful services to researchers. Through an examination of data services and repository options available through the UW-Madison libraries, we can share that some of the approaches that have led to successful outcomes have included being responsive to local needs and contexts, collaborating with campus partners to share infrastructure, and building a community of partners with unique expertise to help develop services that are valuable to the institution as a whole.

Bibliography

"Increasing Access to the Results of Federally Funded Scientific Research." President Obama White House Archives. February 22, 2013. https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf.

"Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." White House. August 25, 2022. https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

"Vision & History." Research Data Services. N.d. https://researchdata.wisc.edu/our-services-2/vision-history/