Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations

Sophia Lafferty-Hess,[1] William Krenzer,[2] Jenny Ariansen,[3] Jennifer Darragh[4]

Abstract

Over the past few years, the United States has implemented a second round of data management policies, exemplified by the 2023 NIH Data Management and Sharing Policy and 2022 "Nelson Memo." Effectively supporting public access to data and a data sharing culture at an academic research institution requires collaboration across various research support staff and central offices as well as knowledge of the current practices of researchers. Two research support groups at Duke University, the University Libraries (DUL) and the Office of Scientific Integrity (DOSI), have forged a strong working relationship for supporting data management and sharing practices, including an active Teams channel for communication, developing tools collaboratively, delivering trainings, and providing co-consults for data management. To more effectively understand "the state of play" at our institution, DUL and DOSI analyzed data management and sharing plans (DMSPs) submitted to the National Science Foundation (NSF) in 2021. The project team used a modified version of the DART rubric (https://osf.io/qh6ad/) to score DMSPs against required elements in key areas, including types of data; standards for data and metadata; access, sharing, and preservation; limitations on access, distribution, and reuse; and roles and responsibilities. In this paper we will present the key findings from the DMSP assessment project and discuss how, as data management specialists, we can use this information to plan for ongoing education, training, and resource development using a cross-campus collaboration model.

Keywords

data management plans, data sharing, institutional collaboration, data management services, repositories

Introduction

The value of shared data has been widely discussed from its impact on the development of the rapid responses to global health crises (Moorthy et al., 2020) to the use of shared data in creating models and training for artificial intelligence (Bell & Shimron, 2024). Open data sharing is evolving from an aspirational goal to a standard requirement by funders (NIH, 2020) and publishers alike (Naughton & Kernohan, 2016). With these growing requirements, academic institutions are developing and refining services to support researchers' needs to effectively manage their data, write data management and

sharing plans for grant proposals, package data for sharing through online repositories, and address often complex ethical and legal dimensions of data sharing.

There has been an evolution of federal funding policies regarding data management and sharing in the United States over the past 15 years. The National Science Foundation (NSF) was one of the first funding agencies to implement the requirement for a "data management plan" (DMP) as part of all grant proposals in 2011, which was quickly followed by the Office of Science and Technology Policy memo "Increasing Access to the Results of Federally Funded Scientific Research" released in 2013 (Pasek, 2017). Since 2020, there has been a new wave of data management and sharing policies within the US. These new policies began with the release of the National Institutes of Health (NIH) Data Management and Sharing policy in 2021 (NIH, 2020), which went into effect January 25th, 2023, and was followed by the 2022 Office of Science and Technology Policy memo "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research" (e.g., Nelson Memo) (OSTP, 2022).

A noteworthy shift in the NIH policy was the more serious implications for non-compliance, as the data management and sharing plan (DMSP) is now considered a Term and Condition of the award, as well as clearer expectations that an established repository is used when making data publicly available (NIH, 2020). In addition, the US government has defined a set of desirable characteristics for data repositories centering on making data more Findable, Accessible, Interoperable, Reuseable (Wilkinson et al., 2016; National Science and Technology Council, 2022). The shift towards asserting the importance of more formal sharing of data within repositories is also exemplified by other federal policies shifting to using the language "data management **and sharing** plans" versus "data management plans" (NSF, 2023, pp. 15). Moving forward, in this paper when referring to data management plans, we will refer to them as DMSPs.

These types of policies have been one of the motivators for the rise of data management services being available within academic libraries (Tenopir, Sandusky, Allard & Birch, 2014). Because of this, numerous data professionals and researchers over the years have begun examining these data management plans. Bishoff and Johnston (2015) used an "opt-in" approach for creating their sample of plans at the University of Minnesota. Others have gained access to a fuller corpus of DMPs at their institution via a particular funding agency (Van Loon, 2017). A study at Purdue focused their research on examining plans created within their institutional data management planning tool (Green, Cairns, & White, 2019). While many of these projects were based at individual institutions, Parham et al. (2016) developed a standard rubric to assess NSF plans across numerous US-based institutions as part of the DART project. While examining DMPs isn't necessarily a "new area of study"; getting a glimpse into researcher's practices can be particularly valuable for building a shared understanding about the current "state of play" at an institution regarding data management practices.

While many data management services over the years have been centered within academic libraries, other units on campuses, particularly research offices, are increasingly important partners, especially regarding the importance of compliance with funder policies around data management and sharing (Reinhart, 2016). These kinds of partnerships between groups helps provide various forms of data management support and are paramount for avoiding "siloing" individual groups and encouraging a collaborative environment for supporting an institution's research community as a whole. For the past

eight years, Duke University Libraries (DUL) and the Duke Office of Scientific Integrity (DOSI), have forged a strong working relationship around DMSPs, including an active *Teams* channel for communication, developing tools collaboratively, delivering trainings, and providing co-consults for data management.

In 2020, DUL and DOSI began a collaboration on a new project aimed at assessing DMSPs from NSF. The project team included two staff members from DUL and two from DOSI. The two staff from the library both had master's degrees in library and/or information sciences and had worked as data librarians or data management professionals for numerous years. Whereas the staff from DOSI both had research experience, one had managed a large lab and had an MS in Neurobiology, and the other had a PhD in psychology. The team viewed this project as (1) a way to gain a better understanding of current practice, particularly as Duke was working on implementing a new institutional Research Data Policy,[5] and (2) as a way to prepare for the new NIH Data Management and Sharing Policy that had come out. An analysis of the plans would also be a foundation for the development of a strategy for enhancing data management education and areas for continuing collaboration between the two groups that would harness the varying expertise, values, and ideals that motivate the groups.

Methods

In hopes of getting a clear and complete understanding of the planned data management practices of our university's researchers, we aimed to review an entire funding year's worth of DMSPs. To get this, the DOSI team exported from our internal grant tracking system a list of all "New" proposal submissions (therefore excluding supplements, non-competing renewals, and competing renewals) to the NSF that occurred between January 1st, 2021 and December 31st, 2021. While there were 250 total submissions in this period, we limited the sample to those that received funding in order to have a more manageable and meaningful dataset. This query resulted in 81 proposals for review. Of these 81, 10 were excluded because the proposal was mislabeled as new when it was actually a supplement for a previously funded proposal. An additional 11 proposals were removed from the sample because we were unable to locate the proposals, and therefore could not get the DMSPs from them. Lastly, an additional two proposals were excluded from the sample because the DMSPs associated with the funded projects were nearly non-existent and their inclusion in the sample would have skewed the data set. This gave us a final sample size of 58 NSF funded proposals with DMSPs to review from the following colleges: Art & Sciences, Engineering, and Environmental studies.

For the review process, the authors used an adapted version of the DART rubric (Whitmire et al., 2016) developed to assess NSF DMSPs. The project team adapted the rubric to create generalized sections that also aligned conceptually with the new NIH data management and sharing policy to allow for a better general understanding of data management practices (see the Appendix for the full rubric). The primary modifications included having one section of the rubric focused more explicitly on access, sharing, and preservation and another section focused on elements that would limit the access, distribution and reuse of data, as well as adding an oversight, roles, and responsibilities section. Each of the four authors individually ranked each element of the rubric on a 3-point performance level scale (complete/detailed (2); addressed issue, but incomplete (1); did not address (0)). The DMSPs were also given a final score of exemplary, satisfactory, or needs improvement (using the same 3-point scale).

The reviews of the DMSPs were done individually via a REDCap form that was generated for this project. From there, across a total of 15 sessions spanning 13-months (February 2022 to March 2023), the authors met to compare the set of ratings for each DMSP. The goal of these meetings was to come to a consensus on the rating of the DMSP for each individual question. While the group sought consensus, no rater was required to change their score (more details about scores and the rate of change can be found in Table 1).

The comparison process only used a subset of questions from the full rubric that either applied to all the NSF directories or that was identified as having particular importance to data management resource and service assessment at Duke. A subset was also used for the in-depth comparison process as this was highly resource intensive, and the project team's main goal was to gain a better conception of data management practices at a general level to determine areas for service and education enhancements.

Of particular interest to the project team was also whether a plan explicitly listed the name of a data repository within the plan and the mechanism for data sharing, as sharing data within a data repository is being increasingly encouraged and recommended in funder and journal policies (NIH, 2020; PLOS, 2014). Plans could include more than one repository as well as more than one mechanism for data sharing. One project team member then went through a process of coding repository and data sharing questions; these codes were then verified by another team member for correctness. When a repository was named, each repository was assigned a category of either institutional, domain-specific, general, or other. Regarding the mechanism for data sharing, the categories included data repository, project website, journal supplement, article publication, by request, code repository, preprint server, other, or NA.[6]

Results

Over the course of 13 months, four raters reviewed 58 DMSPs from funded NSF projects in 2021 on 10 adapted DART questions, for a total of 2,184 ratings given. During the comparison meetings, the group of raters in sum changed 13 of the scores to N/A, leading to a final count of 2,171 ratings given. With that information, a Fleiss' Kappa test was run to test the level of agreement amongst the four raters in assigning the 3-point categorical level labels (Fleiss, 1971). Overall, the kappa statistics showed a substantial agreement amongst the raters when reviewing these NSF DMSPs, $\kappa = 0.71$, $p < .001$ (Landis & Koch, 1977). While Table 1 gives a breakdown of the kappa statistics for each question, we see that the raters had almost perfect to substantial agreement respectively with their ratings for questions "If the data are deemed to be of a "sensitive" nature, the DMSP describes what protections will be put into place to protect privacy or confidentiality" ($\kappa = 0.86$, $p < .001$), "Provides details on when the data will be made publicly available" ($\kappa = 0.81$, $p < .001$), and "Identifies metadata standards that will be used for the project" ($\kappa = 0.80$, $p < .001$). It is important to note that through the consensus meetings with the raters, approximately 12% of the ratings were changed after the meetings, which would increase our level of agreement. However, given that these meetings were used to discuss our scoring reasoning, and with the fact that no one was required to change their score, we do not see an issue with this.

Table 1: Kappa Score

	Total Ratings	Total Changed	% of Change	Fleiss' Kappa	*p*-value
Overall	2,171*	290	12.23	0.719	< .0001
Describes what types of data will be captured, created, and collected	232	6	2.59	0.487	< .0001
Identifies metadata standards that will be used for the project	232	35	15.09	0.800	< .0001
Describes data formats created or used during the project	232	29	12.50	0.634	< .0001
Provides details on when the data will be made publicly available	232	64	27.59	0.811	< .0001
Describes how the data will be made publicly available	232	13	5.60	0.300	< .0001
Describes plans for archiving and preserving digital data	232	26	11.21	0.534	< .0001
If the data are deemed to be of a "sensitive" nature, the DMP describes what protections will be put into place to protect privacy or confidentiality	232	26	11.21	0.865	< .0001
Describes security measures that will be in place to protect the data from unauthorized access.**	94*	36	38.30	0.773	< .0001
If there are factors that limit the ability to share data (e.g., commercialization or proprietary nature of the data), the plan describes those conditions and describes to whom the data will be made available and under what conditions.	232	55	23.71	0.702	< .0001
Final overall assessment of the DMP	232	0	0	0.366	< .0001

Items are displayed by question order to keep thematic data management questions together and in the order they were answered.

*The total ratings score does not include ratings that were switched in N/A while the change score does

**Question 8 on this list only appeared to reviewers that DID NOT mark question 7 as N/A.

Review Descriptions

From there, we can look at the actual breakdown of how the questions were answered by the raters (for detailed breakdown, see Table 2). Starting with the Overall Assessment score, we see that the mode, the most common response, given by raters when reviewing the DMSPs was 'Satisfactory', making up 48% of the scores. The questions that mode response was 'Complete/Detailed' when reviewing DMSPs were "Describes what types of data will be captured, created, and collected" (71%), "Describes how the data will be made publicly available" (53%), and "Describes data formats created or used during the project" (50%). On the opposite end, the only question that the mode response was 'Did not Address' when reviewing DMSPs was "Identifies metadata standards that will be used for the project" (43%). Taking all of this together, we can see that the raters found the DMSPs to be very good at describing specific information about the data (i.e., types, formats, availability, etc.) but very poor at describing the metadata standards that would be used for the project.

Table 2: Response Counts

	Complete / Detailed (2)	Addressed issue but incomplete (1)	Did not address (0)	N/A (3)	*Mean*
Describes what types of data will be captured, created, and collected	167	60	5	0	1.70
Identifies metadata standards that will be used for the project	43	89	100	0	0.75
Describes data formats created or used during the project	118	82	32	0	1.37
Provides details on when the data will be made publicly available	41	109	82	0	0.82
Describes how the data will be made publicly available	125	105	2	0	1.53
Describes plans for archiving and preserving digital data	111	105	16	0	1.41
If the data are deemed to be of a "sensitive" nature, the DMP describes what protections will be put into place to protect privacy or confidentiality	16	24	9	183	1.14

Describes security measures that will be in place to protect the data from unauthorized access.*	15	24	8	36	1.15
If there are factors that limit the ability to share data (e.g., commercialization or proprietary nature of the data), the plan describes those conditions and describes to whom the data will be made available and under what conditions.	20	46	2	164	1.26
Final overall assessment of the DMP	26	113	93	0	0.71

Items are displayed by question order to keep thematic data management questions together and in the order they were answered.

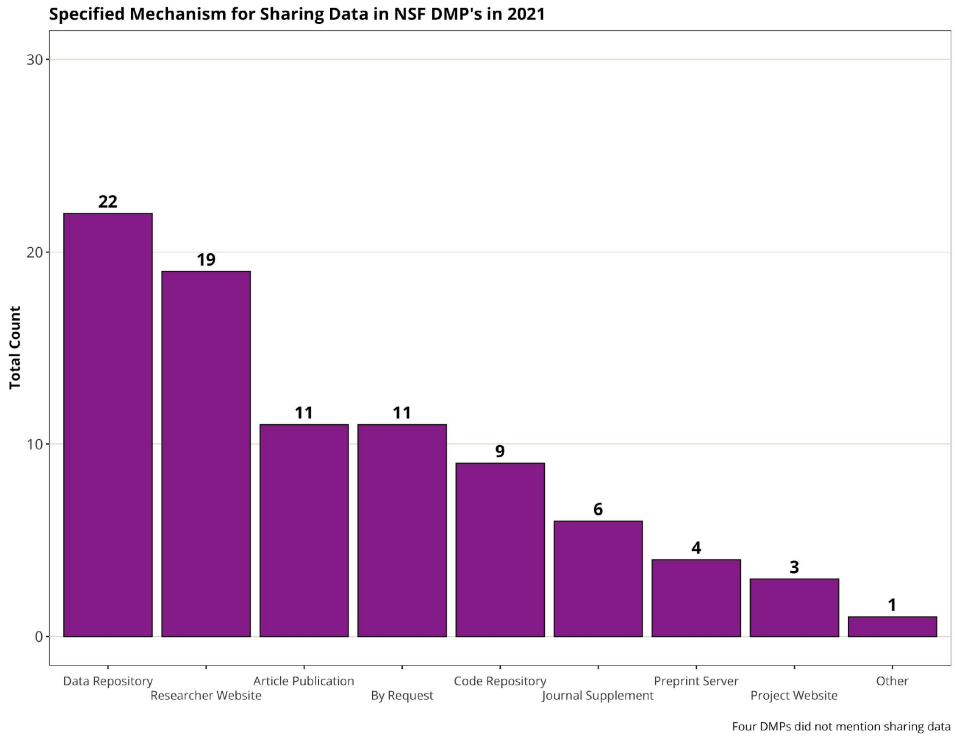
*Question 8 on this regarding security measures list only appeared to reviewers that DID NOT mark question 7 about the sensitive data as N/A.

Data sharing and categories of repositories

Lastly, we reviewed the ways in which DMSPs discussed if, where, and how they would share their data. As a reminder, this data was created by two of the original raters (one did the initial coding of the repository information, while the second verified the information). Additionally, any given DMSP could have named multiple mechanisms/repositories for how or where they would share their data, therefore there was not a 1-to-1 breakdown of DMSPs reviewed and mechanisms/repositories mentioned. But, for our coding, a DMSP could not be coded for the same mechanism/repository multiple times.

Of the 58 DMSPs that were reviewed, a total of 86 different mechanisms for how data would be shared were mentioned (see Figure 1 for full breakdown), with sharing via a data repository as the most common mechanism, appearing in 22 of the DMSPs. Most surprising was that there were four DMSPs that made no specific mention of how or where they would be sharing their data. From there we saw that of the 58 DMSPs reviewed, 56% (33) of them mentioned a type of repository that they would use for sharing their data (Figure 2). Of these 33 DMSPs, there were four types of repositories that were categorized by the coder: General, Institutional, Domain-Specific, and Other. As we can see from Figure 3, Domain-Specific repositories were the most common type of repository, appearing in 57% of the DMSPs, followed by Other (39%), Institutional (21%), and General (12%). Lastly, we identified a total of 25 different repositories that DMSPs mentioned for where they would share their data (Figure 4). The most common repositories mentioned were: Github (27%), Duke Research Data Repository (21%), Sequence Read Archive (12%), GenBank (12%), and arxiv.org (12%). The other remaining 20 repositories mentioned made up a total of 28 responses across the 33 DMSPs that mentioned repositories (see the Appendix for full list of repositories mentioned).

Figure 1. Sharing mechanism identified in plans. Plans could indicate more than one mechanism for sharing.



 Lafferty-Hess, Sophia, Krenzer, William, Ariansen, Jenny & Darragh, Jennifer (2025) Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations, IASSIST Quarterly 49(4), pp. 1-19. DOI: https://doi.org/10.29173/iq1168

Figure 2. Count of plans that named a repository or repositories

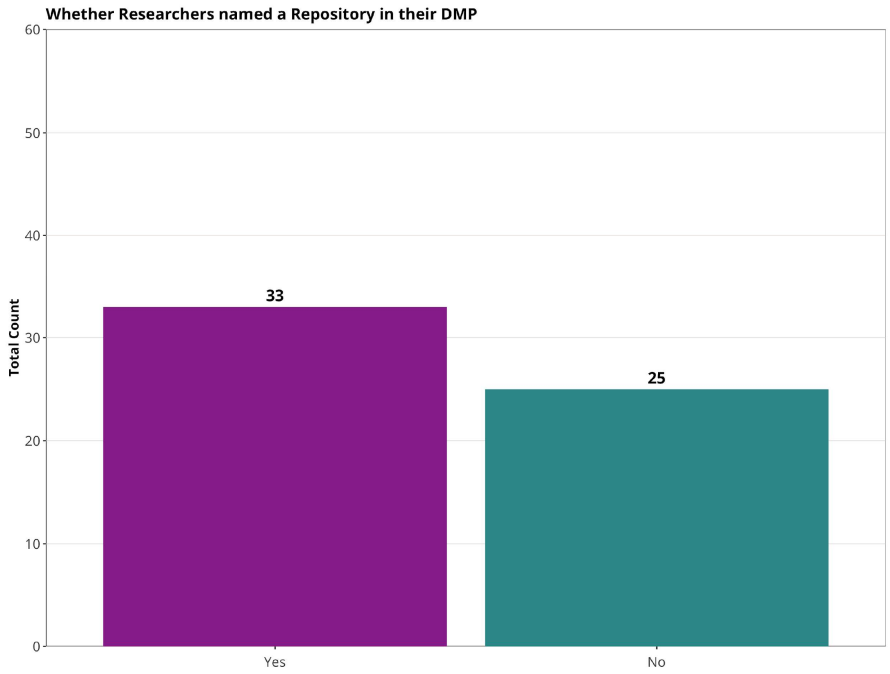
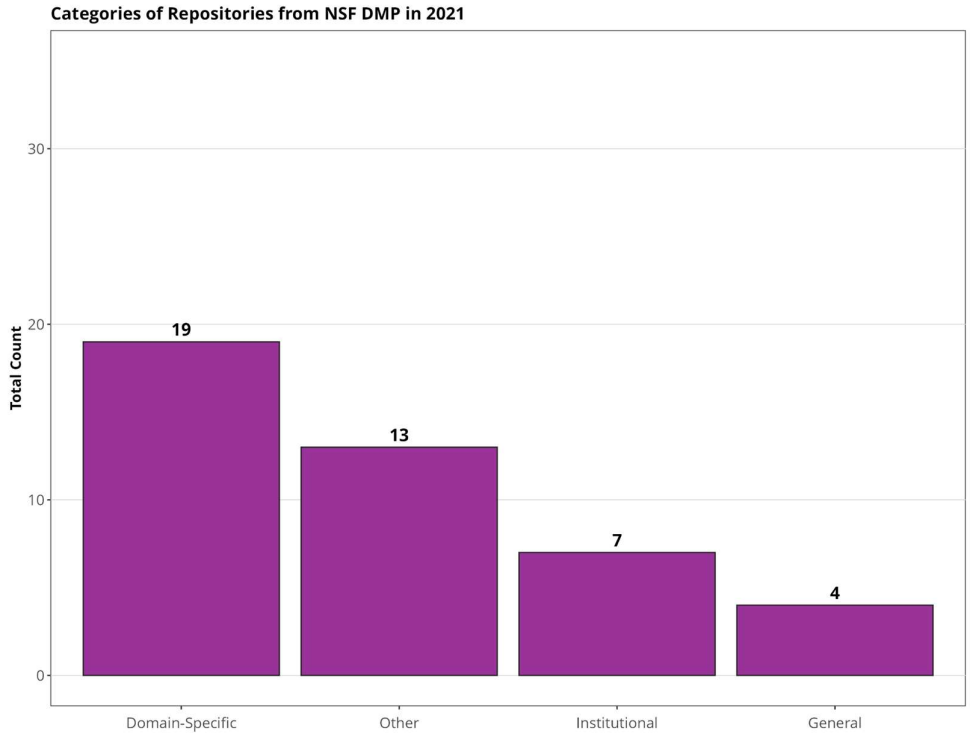
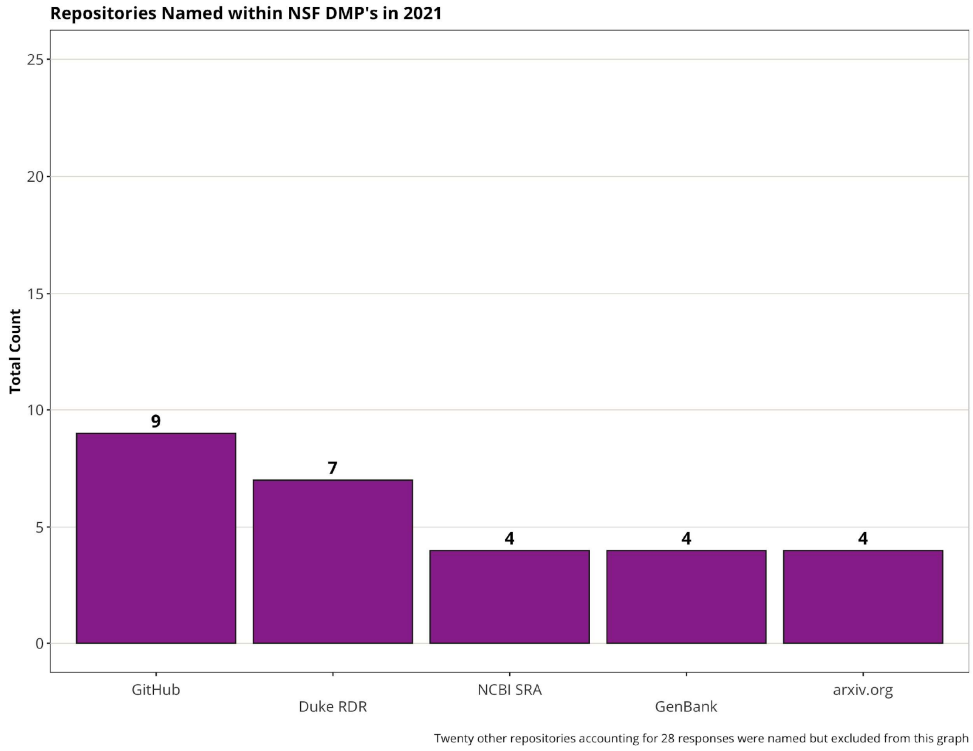


Figure 3. Categories of repositories listed in plans



 Lafferty-Hess, Sophia, Krenzer, William, Ariansen, Jenny & Darragh, Jennifer (2025) Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations, IASSIST Quarterly 49(4), pp. 1-19. DOI: https://doi.org/10.29173/iq1168

Figure 4. Count of repositories named in plans



Discussion

Assessing the content and quality of DM(S)Ps is subjective

While our kappa-level agreement overall showed significant agreement for plans, during the process of comparing coding across our key rubric elements, it became clear that while some metrics seem relatively straight forward, being able to parse the quality or presence of certain content within a DMSP is open to interpretation. The conversations the team had regarding why each of us scored an element the way we did illuminated both how plans require very close reading and re-reading to identify certain content and the variability in interpretation of certain terms. For instance, what do we mean by "metadata standards"? Or what is the quality ranking for a plan that mentions that data will be shared as part of an article? During our re-coding process, the project team also added our own notes to the rubric to clarify some edge cases to aid in better consistency as we moved forward.

While not necessarily surprising, this variability suggests that researchers may receive different advice/feedback when talking to different data management professionals. Likewise, peer reviewers, in the case of NSF, and Program Officers, in the case of NIH, that review the completeness of a plan may come to different conclusions regarding the quality of a plan. This variability also raises questions about how we might add more consistency into the DMSP review process. Is additional training for plan reviewers needed to aid in better consistency and to build a shared understanding of what "makes a good plan"? Could new AI tools assist in performing more machine-actionable assessments of DMSPs? For instance, a recent project has been exploring the potential for developing machine-

 Lafferty-Hess, Sophia, Krenzer, William, Ariansen, Jenny & Darragh, Jennifer (2025) Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations, IASSIST Quarterly 49(4), pp. 1-19. DOI: https://doi.org/10.29173/iq1168

actionable DMSPs more programmatically and this is an area that is worth watching closely as a community (Praetzellis, Gracy, & Taylor, 2025).

Completeness in plans and areas for increasing education and training

Regarding the actual completeness of the answers to specific elements in a plan, the question where we found researchers had the most complete and detailed answers (mean 1.70) was related to the types of data they captured, created, and collected. Whereas the question with the lowest completeness related to the plan identifying metadata standards (mean 0.75). This is not surprising that researchers generally understand the types of data they will be using in their project, whereas metadata standards are more obtuse or in some cases may not be available for a certain data type/discipline. However, this finding also suggests that more targeted education on what are metadata standards could aid more researchers in answering metadata related questions in DMSPs. These forms of education regarding metadata standards may be included within general data management and sharing plan trainings or integrated into more concept specific trainings. For instance, DUL has a Responsible Conduct of Research training aimed at graduate students on the topic of data publishing that first demonstrates the use of metadata standards within several disciplinary repositories and then has participants find a repository based on a case study and rate the FAIRness of repositories asking them among other metrics to assess if and how a repository assigns standardized metadata. This approach goes beyond telling researchers what a metadata standard is and encourages them to engage with them in the real world.

Given the primary purpose of DMSPs is increasingly being communicated in new regulatory documents (NIH Data Management and Sharing Plan) as facilitating public access to research data, it was noteworthy that most plans did address in some way how data will be made publicly available and had the second highest mean for completeness (1.53). However, this is also an area where our team had the lowest percent of change in our scoring (5.6%) and the lowest kappa value (0.300). Meaning this was a question where individual reviewers felt more strongly about their original score and were less likely to modify a score after our group conversation. This also suggests that team members interpreted different forms of making data publicly available (for instance using a data repository vs. making data available as a supplement within a journal article) with higher rates of variation.

It is also noteworthy, although unsurprising given the scope of many NSF sponsored projects (versus what we might see for NIH), that most plans did not have data that would be deemed sensitive or had factors that would limit the ability to share data. These questions were of particular interest to the project team as the risk factors to the institution when dealing with these forms of data increases substantially when sharing data. However, the lower percentage of plans that had complete answers to these questions, recommends this as an area for increasing outreach and resources to ensure that researchers are limiting exposure of sensitive data.

Finally the overall assessment scores for the DMSPs (mean 0.71), suggests that more general education on writing quality DMSPs will benefit the broader community. This is an area where the libraries and the Office of Scientific Integrity have partnered extensively and invested significant time and resources for the implementation of the new NIH Data Management and Sharing Policy. This has

included offering a presentation to all NIH funded departments prior to the implementation of the NIH policy to raise awareness of DMSP review services, implementing new mechanisms to request support via our research portal MyResearchHome, and providing a standard training every semester in partnership with our Responsible Conduct of Research program for faculty and staff on "Meeting Data Management and Sharing Requirements" as well as offering many sessions to individual departments, labs, and groups. Within these trainings, while generally covering elements of DMSPs and the new NIH requirements, we have also integrated information on key terminology that we have seen more limited understanding from this research including metadata standards and what a repository is (and is not). DUL and DOSI have also been discussing how we might partner to reach NSF researchers when the new NSF public access policy goes into effect, albeit that timeline is more tenuous now with the current shifting political landscape within the United States.

Data sharing practices and use of repositories

Through the analysis of questions related to mechanisms for data sharing and the use of repositories, the project team found that over half of the plans (56%) did name the use of a "repository" for data sharing and the most common category of repository used were domain-specific repositories (60%). These findings suggest that some Duke researchers are already well-versed in potential domain-specific repositories in their fields of study, which while not surprising is encouraging. However, it is also important to note that the project team construed "repository" broadly to also include code repositories (e.g., GitHub) and preprints such as arxiv.org. These forms of "repositories" do not provide the necessary features to be considered "data repositories" that meet the desirable characteristics of data repositories or many FAIR principles. Pointing to another area for increased outreach and education to help researchers understand the benefits and limitations of certain platforms and what a formal data repository can provide them.

Likewise, when considering the overall mechanism for data sharing, while "data repository" was the most commonly mentioned (22 plans), the next three highest mechanisms were "researcher website" (19 plans), "article publication" (11 plans), and "by request" (11 plans). These findings also point to the importance of continuing to educate on the value add of using a publicly accessible repository and that there is still much work to do to enhance a culture of public access to data within the United States. To further help Duke researchers navigate the complexities of data sharing, DUL and DOSI developed a data sharing guidance page[7] within MyResearchPath that surfaces regulations, offices to consult, and Duke supported repositories and resources.

Finally, the team members were also particularly interested in the prevalence of the use of the Duke Research Data Repository (RDR) within DMSPs. While we were pleased to see that the RDR was named the second highest (7 plans); however, the fact that GitHub was named the most (9 plans), suggests that additional outreach on the availability and features of our institutional data repository is an area to invest additional resources. DUL and DOSI have partnered over the last two years to extend this outreach including partnering with the Associate Vice President for Research and Innovation to provide presentations to departments and cross-promote the availability of the RDR in newsletters and training.

The value of cross-unit collaborations

This project not only helped the data management specialists at Duke better understand researcher DMSP practices but also was a valuable mechanism for continuing to build a collaborative relationship between DUL and DOSI staff. Working on a project with a shared goal allows units to learn from each other, build a shared understanding, and establish trust, thereby minimizing silos. This project also would not have been possible if undertaken solely by the library as the access to the DMSPs was only possible because the DOSI staff have permissions to access the Duke grants system.

The project team also benefited from the varying expertise and experiences that the cross-unit collaboration brought to bear with the library team members having deeper experience in general best practices in data management, while the DOSI staff had more direct research experience. This allowed the team to have a theoretical/conceptual perspective as well as being grounded in real-world research experience. This can also provide more explanation to the different lenses different team members may have brought to their scoring and the value our conversations around scoring had on expanding our shared understanding of specific terms and practices.

Building on this project, this cross-functional team has engaged in ongoing collaborations including education initiatives, co-consultations, outreach for the Duke Research Data Repository (RDR), and developing a new workflow to assess the appropriateness of NIH DMSPs usage of the RDR as well as promote data sharing resources and guidance to funded projects. The relationships built through the NSF project and other initiatives will be maintained by continuing to share experiences and expertise and working together towards a shared goal to support Duke researchers needing to comply with funder and journal data sharing mandates.

Conclusion

As data management and sharing become increasingly a shared responsibility across various institutional research support offices, the importance of collaboration cannot be understated. At Duke, the project to understand the "state of play" around data management by analyzing a set of NSF DMSPs not only benefited our knowledge but also our relationships. While we did not approach this project with set hypotheses and viewed this research as more exploratory, the project has led us to identify areas for increased education and outreach, particularly in regard to metadata standards, key elements of DMSPs and the interpretation of terminology, resources for the appropriate handling of sensitive data, the purpose and importance of data repositories, and the availability of our institutional data repository. The comparison process also highlighted the complexity inherent in coding qualitative DMSPs in areas/topics that are open to interpretation. We have also found that the process itself helped the team come to a better shared understanding of various terms and practices. Overall, this type of project can create a framework to conceptualize data management services within an institution as we work as a team to advance a data sharing culture at Duke.

References list

Bell, L. C., & Shimron, E. (2024). Sharing data is essential for the future of AI in medical imaging. *Radiology: Artificial Intelligence*, 6(1), e230337. https://doi.org/10.1148/ryai.230337

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5), 378–382. https://doi.org/10.1037/h0031619

Green, P., Cairns, A., & White, H. (2019). Evaluating data management plans—are they good and are they effective? *Proceedings of the IATUL Conferences*. https://docs.lib.purdue.edu/iatul/2019/fair/3

Lafferty-Hess, S., Krenzer, W., Ariansen, J., Darragh, J. (2025). Data and code from: Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations. Duke Research Data Repository. https://doi.org/10.7924/r4668q866

Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33(1), 159–174. https://doi.org/10.2307/2529310

Moorthy, V., Henao Restrepo, A. M., Preziosi, M. P., & Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19). *Bulletin of the World Health Organization*, 98(3), 150. https://doi.org/10.2471/BLT.20.251561

National Institutes of Health (NIH). (2020). *Final NIH policy for data management and sharing*. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

National Science and Technology Council. (2022). *Desirable characteristics of data repositories for federally funded research*. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf

National Science Foundation (2023). *Public access plan 2.0*. https://nsf-gov-resources.nsf.gov/pubs/2023/nsf23104/nsf23104.pdf

Office of Science and Technology Policy. (2022). *Ensuring free, immediate, and equitable access to federally funded research* [Memo]. https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf

Naughton, L., & Kernohan, D. (2016). Making sense of journal research data policies. *Insights* 29 (1): 84–89. http://doi.org/10.1629/uksg.284

Parham, S. W., Carlson, J., Hswe, P., Westra, B., & Whitmire, A. (2016). Using data management plans to explore variability in research data management practices across domains. *International Journal of Digital Curation*, 11(1), 53–67. https://doi.org/10.2218/ijdc.v11i1.423

Pasek, J. E. (2017). Historical development and key issues of data management plan requirements for national science foundation grants: a review. *Issues in Science and Technology Librarianship*, 87. https://doi.org/10.29173/istl1709

PLOS One. (2014). Data availability policy. https://journals.plos.org/plosone/s/data-availability

Praetzellis, M., Grady, B., & Taylor, S. (2025). Piloting maDMPs for streamlined research data management workflows. International Digital Curation Conference, The Hague, Netherlands. https://doi.org/10.5281/ZENODO.14969693

Rinehart, A. (2016). Finding the connection: Research data management and the office of research. *Bulletin of the Association for Information Science and Technology,* 43(1), 28–30. https://doi.org/10.1002/bul2.2016.1720430107

Tenopir, C., Sandusky, R. J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research*, 36(2), 84–90. https://doi.org/10.1016/j.lisr.2013.11.003

Van Loon, J. E., Akers, K. G., Hudson, C., & Sarkozy, A. (2017). Quality evaluation of data management plans at a research university. *IFLA Journal*, 43(1), 98–104. https://doi.org/10.1177/0340035216682041

Whitmire, A. L., Carlson, J., Westra, B., Hswe, P., & Parham, S. (2016). The DART Project: Using data management plans as a research tool. Open Science Framework. http://osf.io/kh2y6

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Endnotes

[1] Sophia Lafferty-Hess is a Senior Research Data Management Consultant at Duke University Libraries. She can be reached by email: sophia.lafferty.hess@duke.edu

[2] William Krenzer is a Research Project Manager at Duke Office of Scientific Integrity. He can be reached by email: william.krenzer@duke.edu

[3] Jenny Ariansen is the Director of Research Integrity at Duke University. They can be reached by email: ja267@duke.edu

[4] Jennifer Darragh is a Senior Research Data Management Consultant at Duke University Libraries. She can be reached by email: jennifer.darragh@duke.edu

[5] https://policies.provost.duke.edu/docs/chapter-5-research-data

[6] Data and code are publicly available at: https://doi.org/10.7924/r4668q866

[7] https://myresearchpath.duke.edu/topics/guidance-sharing-research-data

Appendix

Complete List of Repositories Coded

1: GitHub 9
2: Duke Research Data Repository 7
3: GenBank 4
4: NCBI Sequence Read Archive 4
5: arxiv.org 4
6: BCO-DMO 2
7: CUASHI 2
8: Dryad 2
9: LTER Data system 2
10: MaterialsMine 2
11: MorphoSource 2
12: NCBI GEO 2
13: NSF PASTA 2
14: BioModels 1
15: CRAWDAD Collection at IEEE-Dataport 1
16: Cambridge Structural Database 1
17: DMREF HybriD3 Materials Database 1
18: EcoSIS 1
19: Figshare 1
20: Lepbase 1
21: Materials Cloud 1
22: MetaboLights 1
23: OCB Database 1
24: OSF 1
25: StreamPULSE 1

Scoresheet for assessment of Duke Data Management Plans

The purpose of this scoresheet is to help Duke research support staff to generally assess data management plans completeness and alignment with best practices for data sharing as a mechanism to more fully understand current research practices, compliance, areas for ongoing education, training, and resource development. This rubric was adapted from Whitmire et al. (2017), which had an original purpose to assess NSF data management plans and has been generalized and harmonized with new NIH requirements to allow for application across funding agencies.

Given that not all funders require the same elements for all plans, potential criteria that may be directorate specific within NSF or prioritized by new NIH requirements are noted with an asterisk (*) or plus(+) respectively.

Only highlighted elements will be scored for the essential elements of DMPs. Any NAs will not be counted when calculating the final score.

Section 1: Types of data produced and tools/software used

	Performance Criteria	**Complete/detailed (2)**	**Addressed issue, but incomplete (1)**	**Did not address (0)**
1.1	*Describes what **types** of data will be captured, created, and collected*			
1.2*	*Describes how data will be **collected, captured, or created***			
1.3*	*Identifies **how much** data (volume) will be produced*			
1.4+	*Describes **what data** will be shared from the project*			
1.5+	*Describes **tools or software** needed to access the data and how they will be made available*			

Section 2: Standards for data and metadata

	Performance Criteria	**Complete/detailed (2)**	**Addressed issue, but incomplete (1)**	**Did not address (0)**
2.1	*Identifies **metadata standards** and/or metadata formats that will be used for the project*			
2.2	*Describes **data formats** created or used during project*			
2.3+	*Describes **documentation** that will be made available alongside the data*			

Section 3: Access, sharing, and preservation

 Lafferty-Hess, Sophia, Krenzer, William, Ariansen, Jenny & Darragh, Jennifer (2025) Assessing data management and sharing plans: The "state of play" at Duke and opportunities for cross-campus collaborations, IASSIST Quarterly 49(4), pp. 1-19. DOI: https://doi.org/10.29173/iq1168

	Performance Criteria	Complete/detailed (2)	Addressed issue, but incomplete (1)	Did not address (0)
3.1	*Provides details on **when** the data will be made publicly available*			
3.2	*Describes **how** the data will be made publicly available*			
3.3.	*Provides the **name of the repository** where the data will be archived and made available* ***(If 2 or 1 – drop-down to provide free-text for name of repository)***			
	Name of repository listed in the DMP			
3.4.	*What is the mechanism indicated for sharing data? (free text)*			
3.5*	~~*Describes **how long** the data will be retained and made available to people outside of the project*~~	Removed from study instrument prior to analysis		
3.6	*Describes **plans** for archiving and preserving **digital** data*			
3.7*	*Describes **plans** for archiving and preserving **physical** data*			
3.8.	*What is the mechanism indicated for archiving data? (free text)*			
3.9*	*Identified timeframe for how long data will be archived?*			
3.10*	*Describes the policies in place governing the use and **reuse** (i.e., redistribution, derivatives) of the data (i.e., relevant licenses, Terms of Use)*			
3.11*	*Describes data **types or formats** that will be used for **making data available***			

Section 4: Limitations on access, distribution, and reuse

	Performance Criteria	Complete/detailed (2)	Addressed issue, but incomplete (1)	Did not address (0)
4.1	***If** the data are deemed to be of a "**sensitive**" nature, describes what protections will be put into place to protect **privacy** or **confidentiality** of research subjects*			
4.2	*Describes **security** measures that will be in place to protect the data from unauthorized access*			
4.3	*If there are **factors that limit the ability to share data**, e.g. commercialization or proprietary nature of the data, plan describes those conditions and describes to whom the data will be made available and under what conditions*			

4.4	*Describes what **intellectual property** rights to the data and supporting materials will be given to the public and which will be retained by project personnel (if any)*			

5. Oversight, roles, and responsibilities

	Performance Criteria	**Complete/detailed (2)**	**Addressed issue, but incomplete (1)**	**Did not address (0)**
5.1	*Describes the **roles and responsibilities** of parties involved in the project*			
5.2[+]	*Describes how **oversight and compliance** of the plan with be monitored and managed*			

Final Assessment

Total Score:	*Final Assessment:* *☐ Exemplary ( )* *☐ Satisfactory ( )* *☐ Needs Improvement ( )*	*Notes:*

This scoresheet was adapted from:
Amanda L. Whitmire, Jake Carlson, Patricia M. Hswe, Susan Wells Parham, Elizabeth Rolando, & Brian Westra Version: 1.0 | Date: 23 March 2017 | Contact: Amanda Whitmire at thalassa@stanford.edu

This scoresheet is intended to be used with the NSF DMP rubric developed as part of this project. See Whitmire, A. L., Carlson, J., Westra, B., Hswe, P., & Parham, S. (2017, March 24). Rubric & related files. Retrieved from http://osf.io/qh6ad

**Indicates that the performance criterion is NSF directorate or division-specific*
[+] Indicates new element added to rubric to map to NIH elements or other questions of interest to Duke