




A Tiered Model for Data Management, Curation, and Sharing Support in Grant Proposals and Budgets

Andrew Johnson, University of Colorado Boulder, Boulder, CO, US, andrew.m.johnson@colorado.edu 

Abstract

Recent trends in research funder policies and guidelines increasingly point to requirements for more detailed plans for data management, curation, and sharing with an emphasis on including allowable costs in grant budgets to support these activities as well. As a result, libraries and other collaborating units at academic institutions have an opportunity to develop streamlined models that allow researchers to incorporate institutional resources for data management, curation, and sharing into their grant proposals along with costs to support these resources. While the current literature contains examples and discussions of general funding approaches to activities like data curation and long-term preservation, little has been written on formal models or strategies for incorporating data management and curation support into the grant proposal and budget process itself. This article describes the development and implementation of a tiered grant support model at the University of Colorado Boulder that provides three levels of services, infrastructure, and expertise to researchers who are looking to incorporate support and costs for data management, curation, and sharing into their grant proposals. The article also presents lessons learned since the soft launch of the grant support model as well as key challenges that have emerged. This model may serve as an example for other institutions to use or adapt to continue to help researchers meet funder requirements while simultaneously increasing the visibility of existing data management and curation resources and identifying potential new revenue or cost recovery streams.

Received: May 31, 2023 **Accepted:** September 27, 2023 **Published:** November 30, 2023

Competing Interests: The author declares that they have no competing interests.

The *Journal of eScience Librarianship* is a peer-reviewed open access journal. © 2023 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

See <https://creativecommons.org/licenses/by/4.0>.

∞ OPEN ACCESS

Introduction

With research funders increasingly requiring more detailed plans, including associated costs, for data management and sharing in grant proposals, libraries and other units at academic institutions have an opportunity to develop streamlined models that allow researchers on their campuses to incorporate institutional resources for data management, curation, and sharing more easily into their grants. These models may help raise the profile of existing services and expertise for data management and curation in libraries and other campus units, and grant budgets offer the potential to provide revenue or cost recovery streams that can help to sustain or expand data support and infrastructure. Based on over a decade of experience consulting with researchers, interacting with institutional grant support staff, and developing expertise and infrastructure to support data management, curation, and sharing, the Center for Research Data and Digital Scholarship (CRDDS) at the University of Colorado Boulder (CU Boulder) developed a tiered grant support model to enable researchers to identify resources that can be written into grant proposals with mechanisms for estimating appropriate costs to include in grant budgets as required or encouraged by research funders. This article provides a description of how the model was developed—and currently operates—as well as a preliminary analysis of lessons learned and challenges identified in the first years of its existence. This grant support model may serve as an example for other institutions to use and adapt to their local contexts as they seek to develop sustainable and simplified approaches for researchers attempting to address and budget for data management, curation, and sharing needs in grant proposals.

Background

Support for data management, sharing, and curation, has steadily increased at academic institutions over the past decade in response to evolving research funder requirements, and recent developments at the funder level seem likely to accelerate the further development of models for this support (National Institutes of Health 2023, White House Office of Science and Technology Policy 2022). Often, a variety of campus groups, including libraries, research IT, research administration, and others are involved with providing this type of support (Radecki and Springer 2020). Notably, research funders now allow, encourage, and sometimes require principal investigators (PIs) to include costs associated with data management, sharing, and curation in grant proposal budgets as part of their Data Management Plan or Data Management and Sharing policies (National Institutes of Health 2023, White House Office of Science and Technology Policy 2022). While the literature provides examples and discussion of general issues regarding sustainable funding for data management, curation, and long-term preservation (Bloemers and Montesanti 2020; Bourne, Lorsch, and Green 2015; Goldstein and Ratliff 2010; Johnston et al. 2018), little has been written on formal models or strategies for incorporating data management, sharing, and curation support into the grant proposal and budget process itself.

Institutional context

CRDDS formally launched in 2016 as a collaboration between the campus library and research computing units at CU Boulder, but many of the data management and curation, research cyberinfrastructure, and digital humanities offerings it brought together had existed in some form on campus for five years or more (Knuth et al. 2017). CRDDS also interfaces with research administration, other campus entities involved with research data, and faculty and students from every department across campus. The scope of CRDDS is broader than data management, curation, and sharing as it includes other services and expertise in the areas of research cyberinfrastructure, digital humanities, and open access publishing. In addition, CRDDS provides a robust education and training program across all of its areas of expertise. Currently, CRDDS comprises a total of 16 positions with eight people employed in the library and eight people employed in research computing. Two full-time positions on the library side and one on the research computing side of CRDDS provide primary support for data management, curation, and sharing related to grant proposals. While this article focuses primarily on this data management, curation, and sharing support, the scope of the tiered grant support model covers all services and expertise provided by CRDDS, so discussion will touch on notable examples where researchers combined data management and curation support with other aspects of the larger model (e.g., combining research computing infrastructure with data curation activities).

Tiered Grant Support Model

The CRDDS tiered grant support model comprises three levels: basic support, infrastructure support, and enhanced support (Table 1). These tiers differ somewhat from similar categories identified in the literature in that the type of service or support (e.g., education, consultation, infrastructure) does not solely define each tier (Reznik-Zellen, Adamick, and McGinty 2012). Rather, each tier involves a combination of the type of service or support, the amount of time and/or customization required, whether costs need to be included in grant budgets, and whether those costs are standard or predictable. These tiers were developed based on experience with researchers over many years as well as a desire to make the model as easy as possible for potential users to understand. This involved structuring the tiers in a way that allows users to identify which resources can be included in grant proposals without cost or consultation, which resources can be included as long as a standard cost is also included in the grant budget, and which resources would require a consultation with CRDDS personnel in order to determine appropriate costs and other details. Researchers can choose which individual services they want to utilize from each tier, and they can combine services across tiers as well. For example, one researcher might include multiple services from the basic support tier along with a service from the infrastructure tier while a different researcher might only require enhanced support. The grant support model would enable these use cases in addition to any other combination of services within or across tiers.

Table 1: Grant Support Services for Data Management, Curation, and Sharing by Tier and Cost

Service	Description	Tier	Cost
Data Management and Sharing Plan support	Consultation on and review of draft plans; support for customized DMPTool instance	Basic	Free
Education and training	Consultations, workshops, and group trainings on data management topics	Basic	Free
Public access, archiving, and preservation in institutional repository	Review and curation of all data sets deposited	Basic	Free for data sets up to 500 GB
Data storage	Active storage for large amounts of data	Infrastructure	\$45/TB/year
Public access, archiving, and preservation in institutional repository (large data)	Review and curation of all data sets deposited; integration between repository and large scale storage system	Infrastructure	\$450/TB
Dedicated data management, curation, and sharing support	Personnel included on project team to provide dedicated support to meet project needs for data management, curation, and sharing	Enhanced	Salary/benefits for personnel negotiated on a case-by-case basis

Before making information about this model public, a draft was shared with all members of CRDDS as well as with a group of individuals who support research in various capacities across the institution, including associate deans for research, grant proposal analysts, and other campus faculty, administrators, and staff. The model was revised based on feedback received from these groups, and the final version was then posted to the CRDDS website. This immediately provided a place to direct researchers for more information during consultations, and CRDDS also began conducting informal outreach to help spread the word about the model to interested parties across the institution. The following sections provide additional details about each of the three levels of the tiered support model.

Basic support

The rationale for identifying the services and support to include in the basic tier centered on any resources that researchers could incorporate into a proposal without consulting with CRDDS personnel or including associated costs in grant budgets. Of course, consulting with CRDDS personnel prior to submitting proposals could still be valuable for this tier in order to ensure researchers accurately represent the resources they intend to utilize. In some cases, researchers may also request a letter of support for resources in the basic tier to indicate a formal commitment by CRDDS to provide the applicable support during the performance period of the grant. Services and support included in the basic tier include consultations on data management and sharing plans and approaches, standard (i.e., not customized) education and training offerings on data-related topics, and data curation and long-term preservation in the institutional repository for publicly accessible data sets up to 500 gigabytes. These are all services that are regularly offered to the entire campus population at no cost and with no requirement to be affiliated with a grant-funded project. Two full-time positions in the library primarily support these services with oversight, strategy, and policy guidance provided by the head of the unit in which these positions reside. Small portions of positions that support the administration and development of the institutional repository also contribute to the services for providing public access and long-term preservation for data in the institutional repository.

Infrastructure support

The infrastructure support tier includes services for storage, public access, curation, and long-term preservation of large amounts of research data. For data sets over 500 gigabytes, the CU Boulder institutional repository charges a deposit fee per terabyte (rounded up to the nearest terabyte). This fee is based on recovering at least the unsubsidized per terabyte cost of ten years of large-scale active data storage with policies in place for evaluating the possibility of deaccessioning or moving data to offline storage after ten years in cases where data is used infrequently or is determined to be of low or no remaining research value (among other possible factors). The infrastructure tier also includes large-scale active data storage for large amounts of data collected during the grant project that may or may not need to be shared and preserved over the long-term. This storage cost is partially subsidized at an annual per terabyte rate. Both of these services are available at any time to all campus researchers using the same fee structures regardless of whether researchers have included these costs in a grant budget; however, by including these resources in this tier of the grant support model, researchers are able to estimate appropriate costs for long-term data access and preservation and active data storage to include in their grant proposal budgets. As with the basic tier, it is still valuable to consult with CRDDS personnel when preparing these budget estimates, but this tier at least gives researchers the ability to make cost estimates independently if needed (e.g., to meet a proposal deadline). While developing cost models for data storage and long-term access and preservation of data (in this case only for large data) is one of the more straightforward ways to account for data management, curation, and

sharing costs in grant budgets, the ability or desire to charge such fees will vary by institution. Thus, other institutions might not require something equivalent to the infrastructure support tier, but institutions that are interested in charging for storage or repository costs might find the rationale for how CRDDS developed these fee structures of value as a model. The infrastructure tier is supported by the same personnel involved with the basic tier services in addition to one member from the research computing side of CRDDS who facilitates use of the large-scale storage system, including integration with the institutional repository.

Enhanced support

The third tier, enhanced support, covers all involvement of CRDDS personnel and resources beyond what is covered by the basic and infrastructure tiers. This type of support often involves salary costs needing to be written into grant proposal budgets, but these costs typically need to be negotiated between researchers and CRDDS personnel in advance. The most common example of this type of support is including CRDDS personnel as part of the grant team with responsibilities for carrying out active data management and curation beyond the standard end-of-lifecycle curation provided to all data sets in the institutional repository. This may involve developing customized data management and curation workflows, actively creating documentation for data throughout the grant project, preparing data sets for deposit in repositories other than the institutional repository, and additional consulting activities beyond what is provided in the basic tier. The amount of time and effort required to do this for any particular project varies widely based on the amount and type of data expected to be produced, the project goals, the budget that researchers are able to dedicate to data management and curation, and the available bandwidth of CRDDS personnel. All of these factors must be discussed as CRDDS personnel and the researchers negotiate the appropriate amount of salary and benefits to include in the grant budget. In order to account for the time required for these discussions, researchers are required to initiate contact with CRDDS at least one month before their proposal deadline. Currently, there are three positions from the library side of CRDDS who are available to have salary and benefits written into proposals to support data management, curation, and sharing. These include the two primary positions in the library that support the basic and infrastructure tier services as well as the head of the unit in which those positions reside. To date, PIs have included portions of salary and benefits from all three of these positions in current and pending awards as discussed further in the next section.

Lessons Learned

Dating back to its soft launch several years ago, the grant support model has been utilized as an available resource to over 100 grant-related consultations with researchers across campus as part of the basic support tier. Approximately two dozen of these consultations also involved researchers expressing interest in including infrastructure tier costs in their proposals to support data storage and/or long-term data access

and preservation for large amounts of research data. At present, PIs from six separate grant projects have included salary and benefits for CRDDS personnel in proposal budgets to cover enhanced tier support for data management, curation, and sharing. Two of these proposals that involved CRDDS personnel have been funded so far. One of these awarded grants included dedicated support for both research cyberinfrastructure and data curation while the other included enhanced tier support for data curation alongside curriculum development for data literacy education. While it is too early to know if this is the start of a meaningful trend, it is interesting to note that both funded proposals so far included a combination of multiple areas of expertise from the enhanced tier. This could be an early indicator that libraries and other campus units involved in data management and curation might be well-served to take a holistic approach to support for research data needs if they are interested in being included as paid personnel on grant teams. In addition, the awarded grant that combined enhanced data curation and cyberinfrastructure support is a very large interdisciplinary and multi-institution grant (approximately \$25 million budget), which could point to another area of opportunity in targeting or focusing on large proposals for this type of support.

Another lesson learned as the tiered grant support model was being developed and implemented involved how salary and benefit savings from CRDDS personnel included on grants would be handled within the larger institutional context. As a collaboration between the campus library and research computing groups, CRDDS has its own annual budget, but salary and benefit costs typically come out of each unit's separate larger budgets depending on where respective personnel are officially employed. As a result, negotiations with each unit's leadership needed to take place to determine where and how cost savings from salary and benefits included in funded proposals would be returned. As noted in the literature, this can be a difficult issue to navigate for some categories of employees at academic institutions that do not have standard or applicable procedures for course buyouts and other similar forms of compensation (Chaput and Walsh 2023). In the case of CRDDS, the trend to date has been for salary and benefits from grants to return to the annual shared budget to support and expand services, initiatives, and programs. Other institutions may want to consider discussing this issue carefully before developing any grant support model that includes the potential for personnel costs to be included in grant budgets.

Challenges

Several challenges have emerged in implementing the tiered grant support model, particularly regarding the infrastructure and enhanced tiers. The first major challenge is the scalability of enhanced tier support. Because this tier often involves portions of existing positions being written into grant proposals, there is a constantly shifting number of funded proposals that CRDDS is able to support at any given time. In addition, not all proposals end up being awarded, so it can be challenging to navigate how many proposals CRDDS personnel should strategically target in order to maximize potential involvement in awarded grants.

Proposed and awarded grant timelines also overlap, so keeping track of which personnel are involved with which potential and awarded grants creates another logistical difficulty.

Another challenge, which is related to scalability, is turnover in positions involved with supporting proposed or awarded grants. Personnel hires and departures have become increasingly unpredictable at many institutions, particularly since the beginning of the COVID-19 pandemic, and this can create issues for the need to modify or in some cases even back out of commitments to providing enhanced grant support. Because only partial salary and benefits for CRDDS personnel are typically written into grant budgets, turnover in these types of roles can be more problematic in some ways than for a position that is paid fully out of the grant. In the latter case, the PI is often able to hire someone new themselves, whereas it is typically more difficult and time-intensive to initiate and complete the process of rehiring personnel in CRDDS for a variety of reasons. One possible solution to this turnover issue would be for researchers to write in dedicated positions on their grant teams to support data curation, management and sharing, but there is not always a need or available room in the budget for a full-time position like this.

Finally, it would be useful for a variety of reasons to be able to better monitor the content of all grant proposals that include CRDDS resources regardless of whether researchers have consulted with CRDDS personnel. Currently, CRDDS only becomes aware of proposals that include its resources via researchers themselves, so there is no way to know how many total proposals include CRDDS resources for purposes of planning, scalability, and sustainability. In addition, CRDDS has no easy system in place for finding out whether proposals that include its resources, or even its personnel, are ultimately funded. Again, researchers themselves would need to reach out directly to let CRDDS know about this. Researchers have typically done this when CRDDS personnel are involved, and it would be surprising to see that trend change. Thus, the practical implications of this challenge are most relevant for infrastructure tier support where it would be useful to be able to estimate future demands on large-scale data infrastructure as well as the amount of funding that can be expected from awarded grants for budget planning. As the need for data curation infrastructure typically does not arise until later in a grant-funded project, CRDDS may not become aware that grant funds will be used for this support until several years after a successful proposal is submitted to a funder. CRDDS will continue to explore and evaluate possible solutions to all of these challenges going forward.

Conclusion

The CRDDS tiered grant support model provides an example that other academic institutions can replicate or adapt to meet researcher needs as they seek to incorporate support and funding for data management, curation, and sharing into their grant budgets. This can benefit libraries and other units involved in this support by raising the profile of their expertise in data management and curation while potentially bringing

in new sources of revenue and cost recovery. This model is not without its challenges, particularly around scalability and tracking of resources included in awarded proposals, but its relevance will only continue to increase as funders strengthen requirements for including costs associated with data management, curation, and sharing in grant proposal budgets. As a result, it will be important to see how this and other possible grant support models develop in the coming years. This will include monitoring whether early trends indicating a need for holistic research data support beyond just curation, especially for large and interdisciplinary grants, will continue going forward. In addition, CRDDS will need to develop and implement more formal and regular mechanisms for collecting and evaluating user feedback on whether this model is meeting researchers' needs across all three tiers.

Competing Interests

The author declares that they have no competing interests.

References

- Bloemers, Margreet and Annalisa Montesanti. 2020. "The FAIR Funding Model: Providing a Framework for Research Funders to Drive the Transition toward FAIR Data Management and Stewardship Practices." *Data Intelligence* 2(1-2): 171–180. https://doi.org/10.1162/dint_a_00039.
- Bourne, Philip E., Jon R. Lorsch, and Eric D. Green. (2015). "Perspective: Sustaining the Big-data Ecosystem." *Nature* 527: S16–S17. <https://doi.org/10.1038/527S16a>.
- Chaput, Jennifer and Renee Walsh. 2023. "Data Management Librarians Role in a Large Interdisciplinary Scientific Grant for PFAS Remediation: Considerations and Recommendations." *Journal of eScience Librarianship* 12(1): e616. <https://doi.org/10.7191/jeslib.616>.
- Goldstein, Serge J. and Mark Ratliff. 2010. "DataSpace: A Funding and Operational Model for Long-Term Preservation and Sharing of Research Data." *DataSpace*. Last modified August 27, 2010. <https://dataspace.princeton.edu/handle/88435/dsp01w6634361k>.
- Johnston, Lisa R., Jake Carlson, Cynthia Hudson-Vitale, Heidi Imker, Wendy Kozlowski, Robert Olendorf, Claire Stewart et al. 2018. "Data Curation Network: A Cross-Institutional Staffing Model for Curating Research Data." *International Journal of Digital Curation* 13(1): 125–140. <https://doi.org/10.2218/ijdc.v13i1.616>.
- Knuth, Shelley L., Andrew Johnson, Thea Lindquist, Debra Weiss, Deborah Hamrick, Thomas Hauser, and Leslie Reynolds. (2017). "The Center for Research Data and Digital Scholarship at the University of Colorado-Boulder." *Bulletin of the Association for Information Science and Technology* 43(2): 46–48. <https://doi.org/10.1002/bul2.2017.1720430215>.
- National Institutes of Health. 2023. "Final NIH Policy for Data Management and Sharing." *Grants & Funding*. Last modified January 25, 2023. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>.

Radecki, Jane, and Rebecca Springer. 2020. "Research Data Services in US Higher Education." *Ithaka S+R*. Last modified November 18, 2020. <https://doi.org/10.18665/sr.314397>.

Reznik-Zellen, Rebecca C., Jessica Adamick, and Stephen McGinty. 2012. "Tiers of Research Data Support Services." *Journal of eScience Librarianship* 1(1): e1002.
<https://doi.org/10.7191/jeslib.2012.1002>.

White House Office of Science and Technology Policy. 2022. "Ensuring Free, Immediate, and Equitable Access to Federally Funded Research." Last modified August 25, 2022.
<https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>.