



Cross-functional policy development for a Data Repository

Item Type	Article
Authors	Atwood, Thea P.;Jerome, Erin;Kardos, Ann;McGinty, Stephen;Radik, Melanie;Reznik-Zellen, Rebecca
Citation	Atwood T. P. & Jerome E. & Kardos A. & McGinty S. & Radik M. & Reznik-Zellen R., (2021) "Cross-functional policy development for a Data Repository", Journal of Librarianship and Scholarly Communication 9(1). doi: https://doi.org/10.31274/jlsc.12911
DOI	10.31274/jlsc.12911
Rights	UMass Amherst Open Access Policy;Attribution 4.0 International
Download date	2026-05-17 11:17:52
Item License	http://creativecommons.org/licenses/by/4.0/
Link to Item	https://hdl.handle.net/20.500.14394/32287

JLSC

ISSN 2162-3309 | JLSC is published by the Iowa State University Digital Press | <http://jpsc-pub.org>

Volume 9, General Issue (2021)

Cross-Functional Policy Development for the UMass Amherst Data Repository

Thea P Atwood, Erin Jerome, Ann Kardos, Stephen McGinty, Melanie Radik & Rebecca Reznik-Zellen

Atwood, T.P., Jerome, E., Kardos, A., McGinty, S., Radik, M. & Reznik-Zellen, R. (2021). Cross-Functional Policy Development for the UMass Amherst Data Repository. *Journal of Librarianship and Scholarly Communication*, 9(General Issue), eP12911. <https://doi.org/10.31274/jpsc.12911>

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.



© 2021 Atwood, et al. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

PRACTICE ARTICLE

Cross-Functional Policy Development for the UMass Amherst Data Repository

Thea P Atwood

Data Services Librarian, University of Massachusetts Amherst

Erin Jerome

Open Access and Institutional Repository Librarian, University of Massachusetts Amherst

Ann Kardos

Metadata Librarian, University of Massachusetts Amherst

Stephen McGinty

Education Librarian, University of Massachusetts Amherst

Melanie Radik

Science and Engineering Librarian, University of Massachusetts Amherst

Rebecca Reznik-Zellen

Head, Science and Engineering Library, University of Massachusetts Amherst

ABSTRACT

Policy can articulate the scope of work. For repositories that house data, policy can help users manage expectations, especially for individuals who are new to data sharing, or where expectations for sharing data have changed. We cover some of the current literature around the process for writing policy, specifically focusing on policy for data collections and repositories, factors that encouraged us to create a repository policy, our collaborative process for creating the policy, and lessons learned. We hope that others can use our processes to build their own policy that reflects the needs of their campuses and scholars and further moves the needle toward the “Library as Publisher” model.

Received: 12/20/2019 Accepted: 03/25/2021

Correspondence: Thea P Atwood, Data Services Librarian, University of Massachusetts Amherst, tpatwood@umass.edu



© 2021 Atwood, et al. This open access article is distributed under a Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

INTRODUCTION

Data stored in an institutional repository (IR) has several advantages. Not only does locally stored data showcase the outputs of an organization, but repositories—as compared with personal websites—can help bolster views and downloads. The reach of what is in a repository can be improved with a dedicated communications plan or team. A clear line of responsibility for a Twitter account, for example, can have a great impact on marketing all content in a repository (e.g., Peoples et al., 2016). The citation advantage for open data could incentivize researchers to share their research outputs (e.g., Piwowar and Vision, 2013; SPARC Europe, 2017). Some repository platforms report on the metrics of use for datasets, further demonstrating the impact of a researcher’s work. IRs can act as a home for data without a dedicated discipline repository.

Policy for repositories plays an important role in dictating scope and expectations. Although it is beneficial to have a policy in place prior to launching a new initiative, this is not always how services and resources develop, and it may not always be practical. In our case, our collection of data in our IR—ScholarWorks@UMassAmherst (or “ScholarWorks”; scholarworks.umass.edu)—predated our policy. Locally, we refer to this collection as our data repository although it is technically a large collection within our IR. We will refer to it as our data repository throughout the rest of this article. Furthermore, although the data repository exists under the auspices and policies of the larger repository, we found it necessary to create a specific policy for the data repository. In searching, we found little guidance related to the process of writing a policy for data collections or data repositories, although we found a handful of existing policies developed at peer and aspirant institutions.

With the need to establish a policy for the data repository, the University of Massachusetts Amherst (UMass Amherst) Libraries’ Data Working Group collaboratively approached writing a policy for depositing data in it. We shared findings, brainstormed aspects that were most relevant and of highest priority to our campus, and crafted this public-facing statement. And although there is a great deal of discussion around establishing a repository—and some discussion of policy generally, including surveys of professionals and content analyses—we found little guidance on the act of collaboratively writing a policy. As such, we are documenting here our distributed approach to policy writing, the lessons learned, and next steps. We hope to help other libraries and librarians working to write their own policies—for data or other—by offering an explanation of our approach and processes.

We also hope to contribute to an ongoing conversation about standardizing components of repository policy. This discussion could also help other organizations just beginning to explore hosting data at their institution. Because creator retention of the ownership of data is

important, largely because publishers see it as having great value to exploit (Aspesi et al., 2019), we hope that this groundwork further encourages more universities to explore hosting their research data.

LITERATURE REVIEW

IRs have grown in scope—beginning as a hosting platform for previously published content and, over time, growing to include original publications, gray literature, theses and dissertations, conference proceedings, journals, presentations, research data, and more. IRs play an important role in capturing the scholarly output of a campus.

Literature on a library-based data repository policy exists in a few different categories. These include the following: explorations on the scope and content of a policy, including exploring the concepts and purpose of a repository policy (Riddle, 2015), emerging workflows in data management in a repository (Austin et al., 2017), and content analysis of existing policies (Higman & Pinfield, 2015); how policies are developed and who should be included, including the importance of stakeholders in policy development (Erway, 2013; Shearer, 2015; Tenopir et al., 2017; Van Zeeland & Ringersma, 2017; Verhaar et al., 2017) and the development of policies occurring from the “bottom up” (Lee and Stvilia, 2017; Cruz et al., 2019); tying policy to services, including the relationship between data management policy and services (Higman & Pinfield, 2015); the lack of policy and the need for standards, including discussions on the lack of standards (Briney et al., 2015; Austin et al., 2017); and the lack of research data management policy coupled with the lack of strategic development of these policies (Yu, 2017). Unfortunately, these resources did not provide guidance on how to approach writing our own policy, what has worked for others, what challenges they faced, and how their policy changed over time.

Existing frameworks and guides also seemed outside the scope of our purpose: to write a policy for a newly established data repository within a larger, library-managed IR. The Data Audit Framework (Jones et al., 2008), suggested by Anderson (2010) for digital libraries, was too broad—the focus of this framework is on institution-wide data location and responsibility efforts. However, Anderson’s call to ensure that the social sciences and humanities be included when developing repositories is an important piece of capturing the scholarly output of a campus and is included in the future work of the campus and the Data Working Group. Guides such as “Policy-making for Research Data in Repositories: A Guide” (Green et al., 2009) were missing pieces of the policy-writing puzzle; for example, there is little discussion of end-user agreements. Guides may also be too detailed for an external-facing document meant to assist end-users. Our first policy was to be agile, adaptive, and approachable, not comprehensive, immobile, and daunting.

Since writing our policy, some information has been published around frameworks for digital preservation in trusted digital repositories (Lin et al., 2020), and there are calls from the US government around standards for data repositories (e.g., the Office of Science and Technology Policy’s “Request for Public Comment on Draft Desirable Characteristics of Repositories for Managing Data Resulting From Federally Funded Research” and supplemental information to the NIH’s policy for data management and sharing, “Selecting a Repository for Data Resulting from NIH-Supported Research”), but these lack practical guidance on writing a policy for data repositories. These documents are important in informing what could go into such a policy and what funding agencies see as important, so they should be taken into consideration when developing a repository solution.

Because there is little in the way of describing how library data collection or data repository policy is written, what topics are important, and general guidance, we hope to bridge this gap by describing our process here.

BRIEF HISTORY OF THE DATA WORKING GROUP AND DATA IN SCHOLARWORKS@UMASSAMHERST

Brief history of the Data Working Group

First established in 2011, the Data Working Group is a standing committee within the University Libraries. Since its inception, the Data Working Group has provided feedback on data management plans and has offered general education on data management concepts. The group’s charge expanded in 2017, when the data services librarian position was filled. Now—in addition to reviewing data management plans, providing general education, and staying on top of trends on campus—the group provides feedback to the data services librarian. Committee members serve by virtue of their position or their expertise. Thus, the Data Working Group serves as a check on the data services librarian, by offering valuable insight, perspectives from other disciplines, and input from other areas of the libraries and on campus. This makes for a robust group, invested in how data services progresses at the university.

Brief history of data in ScholarWorks@UMassAmherst

UMass Amherst has used bepress’s (bepress.com) Digital Commons as their repository platform for over a decade. Prior to the establishment of the ScholarWorks@UMass Data Repository in October 2017, data and datasets were accepted in the IR on an *ad hoc* basis. Datasets were added to departmental-level collections within ScholarWorks in keeping with the established hierarchical structure of the overall repository (e.g., see https://scholarworks.umass.edu/eco_datasets/).

In order to better showcase the campus's open data, the open access and institutional repository librarian and the data services librarian decided to create a central collection of data, which we refer to as a data repository. This is a distinct collection of data that exists within the larger IR—similar to how electronic dissertations and theses are typically handled.

WRITING THE POLICY

Policy development

Since its establishment, the data repository operated under the auspices of the larger IR policy. The need for a policy for the data repository became pressing when we began to receive requests for data that could not ethically be shared publicly. When discussing the ethics behind why certain data cannot be openly shared, some campus researchers pushed back—they wanted an explanation as to why we were unable to share this data and were dismissive of the indirect identifiers within their dataset. Scholars also wanted the staff of the Libraries to anonymize the data. Although the Data Working Group is composed of several experts, anonymizing data is not within the scope of our work, and we did not want to set an expectation that we were able to take on this task.

In order to provide scholars with a standard point of reference, and to clarify the limitations of our repository infrastructure, the data services librarian suggested that the group write a policy collaboratively. This would leverage our diverse interests and knowledge bases, and it felt like an approachable way to write policy amid a group of individuals with many other roles and responsibilities to fulfill. Thus, the Data Working Group agreed on a policy-writing process that would allow for input from all members of the group. Steps included the following:

- Researching and reviewing other policies
- Writing the policy by dividing up the sections
- Reviewing and approving the collaboratively written policy
- Posting the policy to the repository

We also decided to create a light and agile policy—one that addressed our current issues, not all issues that could potentially arise when depositing data. This was done to limit the scope of the policy and gave us some reassurance that we could adjust the policy as necessary, or when changes in trends or needs arose. We also wanted to keep our policy as straightforward as possible, which included using language that was approachable. With this light, agile, and easily understandable framework in mind, we began crafting our policy.

Research and review other policies

Members of the Data Working Group were each tasked with finding two to four policies related to data deposit in a repository. This range was selected to keep the list of policies manageable and to be cognizant of working group members' time. We recognized that this could pose a limitation to our work, but because we were taking a nimble approach to policy development, we knew we could revise the policy at a later date. We located a total of 11 policies to review. A list of these policies is available in Appendix A.

At a subsequent meeting, we reported on our researched policies, with the goal to have a breadth of concepts to draw from in creating our own repository policy. From the located policies, we selected 11 concepts to explore as part of our own policy. The concepts that we selected were based on mutual agreement that these would be of assistance when working with potential depositors and that the concepts were within the scope of the IR and the goals of the campus.

We agreed that an introduction to the scope of the service (General Statement) laying out information about what type of data we could accept and how others can use it (Data We Collect and How People Can Use Your Material, Terms of Use, Takedown Policy) would help give scholars and researchers a shared understanding of the service. The deposit license and submission agreement were two components that were already in place for other deposits in the data repository and, as such, were included in our initial policy development process. A section on boilerplate language provided a straightforward place in which to host language affiliated with the repository's use—something that is often asked of the Data Working Group during data management plan consultations. A section on registering data gave us an opportunity to fulfill a broader campus desire to capture as much research output from the campus as possible. Although this is not a perfect solution to the complex problem of tracking research outputs, it is a starting place and provides the UMass Amherst Libraries a foundation for growth. We opted for a final section that points to any policies used in the creation of our document as a way to provide credit for the work done by other organizations. Our initial headings and final headings are detailed in Appendix B.

Writing the policy

As a group, we decided to pursue a “divide and conquer” approach to policy writing. Each member of the Data Working Group selected one to two sections to draft for the policy and was in charge of researching that section. Members had approximately one month to work on their sections—the time between standing monthly meetings of the working group. The team knew that we would collectively review the policy at a subsequent meeting.

All work was done in Google Docs.

Reviewing the policy

We used one of our regular monthly meetings to review our individual contributions to the policy. Each section's author described what they wrote, how it related to the larger policy, and how it built on other parts of the policy. This helped the group gain a shared understanding of the policy and, importantly, helped us strengthen the policy by clarifying information and making connections across the whole work.

After the group edited the policy, the chair of the Data Working Group (the data services librarian) reviewed it to ensure a shared voice and clarify any lingering issues. The policy was then sent to the Libraries' copyright and information policy librarian for a critical review. After this review, the language was further refined, topics were clarified, and two sections deemed unnecessary were removed. The "takedown policy" was deemed out of scope, in part due to the complex requirements of copyright law, and "Terms of use" was largely redundant with our "End-User Access Policy." See Appendix B for a comparison of our initial section titles as compared with the final version's section titles.

The policy was sent to the Data Working Group via email one last time for any final edits. Once this step was completed, the policy was finalized and was then formatted in HTML by the data services librarian for posting in the repository.

Posting the policy to the repository

Posting the policy and the supplemental pages to the repository took three separate steps: (1) the "README files for Data and Datasets" templates page (<https://scholarworks.umass.edu/data/guidelines.html>), because other pages needed to link here, done in collaboration with our bepress consultant; (2) the "Policies for Data and Datasets" page (<https://scholarworks.umass.edu/data/policies.html>), which we were able to make available without any additional contact with our bepress consultant; and (3) the submission form, all edits to which required contact with our bepress consultant (Appendices C and D). This page was available last because the templates page and the policies page provided important context. An overview of the submission form, based off of the University of New Hampshire's submission form, is available in Appendix C, and a screenshot of the submission page from the user's view is available in Appendix D. Once all the pages were live, we turned on the "Submit Data" link, allowing for self-deposit of datasets.

Because the group considers this a living document that will be updated as trends change or needs become evident, we were able to limit our concerns around perfection and completeness. We felt confident in posting a policy that is “enough for now.” Including a line in the policy that the policy is subject to change helps give users an expectation that this policy may be modified over time.

The entire process, from project initiation to turning on the “Submit Data” link, took approximately seven months.

Lessons learned

Writing a policy was new to everyone on the team, but we used this as an opportunity to learn together and to strengthen our understanding of the scope of our own data repository. Furthermore, we are grateful to the organizations that created policies for their data repositories, giving us a framework upon which we could build our own policy. Our policy is an amalgamation of several sources, and we strove to write a document that had limited use of jargon and was fairly easy to understand.

Having a librarian knowledgeable about policy was of immense benefit. This guidance meant that our policy met our current needs and did not delve into issues that were beyond the scope of either the repository or the UMass Amherst Libraries (see Appendix B for a comparison of our section titles before and after review). Again, this fell in line with our light and agile approach.

When we found areas where we had some disagreement, we were open to learning and talking about the issue. For example, the selection of a license for data was a point of some discussion. Although our repository does not force the selection of a license (i.e., allowing for the option “None” in the license field), we did want to encourage the use of Creative Commons licenses. Some repositories (e.g., Schaeffer, 2011) only use CC0, the Creative Commons public domain dedication (<https://creativecommons.org/share-your-work/public-domain/cc0>). We elected to allow our users to choose either the Creative Commons Public Domain Dedication (CC0) 1.0 or the Creative Commons Attribution License (CC-BY) 4.0. This choice was made to help provide one of the major carrots of sharing data—citations. A recent report stated that, of the more than 8,000 scholars surveyed, 61% considered full citation as a credit mechanism that would facilitate data sharing (Digital Science, 2019). Because there is no requirement to cite data with a public domain dedication, we decided to offer the Creative Commons Attribution License as a way to mitigate anxieties about data citation and credit for work. As new trends emerge, we can revise our policy.

The data services librarian worked to ensure that we kept to our timeline, but it still took longer than the group had estimated. While the policy-writing component took two months to complete, it took another five months to solicit the policy review, incorporate feedback, and post the policy publicly on the repository. We were open to our colleagues having different priorities than our own and were understanding of their priorities. We also note that working with a third-party vendor, as we do in working with our bepress consultant, was its own bottleneck that slowed our process down. We do not anticipate future revisions of the policy to be as onerous as our initial set of revisions. We plan to review the policy on a yearly basis, as well as in response to a significant issue. This aligns with our iterative approach to policy writing.

Finally, it can be enticing to write a policy that covers all possible situations facing a data depositor. However, to do so is often not realistic: it can be challenging to try and anticipate all scenarios for which you would need a policy. The group was therefore mindful of scope creep—which was also kept in check since the group has a great degree of trust, and we kept to our one to two sections each. However, we were not entirely able to mitigate our scope creep, as the section on “Registering Data” was not included in the initial review of policy topics.

CONCLUSION

In reporting on the current state of the literature, our processes, and our outcomes, the hope is to provide an example for other organizations to follow when writing their own policy. From our own lessons, we suggest that others writing a policy be open-minded about the experience, be generous with their peers and experience levels, become comfortable with a policy that covers “enough for now,” and be mindful of scope creep. If there is someone in the organization who can provide policy-writing expertise, that is of immense benefit. Having someone in a leadership or project management role can help mitigate some challenges that arise when working in a group. We also suggest that those coming to the policy-writing process remember that this is not an exact science; the policy will depend on the scope of the organization.

Relatedly, there may be room for conversations around standardizing college and university data collection or data repository policies, as called for by others (Briney et al., 2015; Austin et al., 2017). Having an example policy may alleviate some of the start-up costs associated with starting a repository, demonstrate to leadership that there are successful examples, or demonstrate that there are important opportunities.

We have identified several areas of improvement for our data repository, including marketing the data repository and the policy (especially with our stakeholders who are part of the research data pipeline), scaling the workflow for self-deposit and mediating datasets, establishing a

review cycle for the policy, improving the language of the policy aligned with the principles of reading on the web (Felder, 2011), and working in concert with other campus offices and entities to develop robust policies with clear lines of responsibility (McCready & Molls, 2018; Patterton et al., 2018).

Even though we have not yet marketed this service, of ten datasets deposited in fiscal year 2020, four were self-deposited. This indicates some desire for a campus option for data deposit.

Finally, we see a related goal of the data repository being one where we help researchers retain rights to their data. This is important in the face of large publishers working to own the entire lifecycle of data—from project conception to dissemination—as seen with publishers such as Elsevier offering solutions for research administrators and project start-up (Pure), capturing how scholars access and use articles (Mendeley), publishing results and tracking citations (ScienceDirect, Scopus), and hosting data (bepress, of which the authors are mindful) (see also the SPARC Landscape Analysis [Aspesi et al., 2019]). We are at a critical point to help maintain a culture of openness with data, thus fulfilling the promise of research to enrich the lives of all, not just those with expansive budgets. We hope that our work helps provide insight into one more tool for colleges and universities to explore and helps others take one more step toward the “Library as Publisher” model (e.g., Lippencott, 2017).

REFERENCES

- Anderson, I. (2010). Digital libraries for the arts and social sciences. In M. Collier (Ed.), *Business planning for digital libraries: International approaches* (pp. 45–55). Leuven University Press. <https://doi.org/10.2307/j.ctt9qdz2g>
- Aspesi, C., Allen, N., Crow, R., Daugherty, S., Joseph, H., McArthur, J., & Shockey, N. (2019, March 29). SPARC landscape analysis. <https://sparcopen.org/our-work/landscape-analysis/>
- Austin, C. C., Bloom, T., Dallmeier-Tiessen, S., Khodiyar, V. K., Murphy, F., Nurnberger, A., Raymond, L., Stockhause, M., Tedds, J., Cardigan, M., & Whyte, A. (2017). Key components of data publishing: Using current best practices to develop a reference model for data publishing. *International Journal on Digital Libraries*, 18(2), 77–92.
- Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication*, 3(2), 1–25. <https://doi.org/10.7710/2162-3309.1232>
- Cruz, M., Dintzner, N., Dunning, A., van der Kuil, A., Plomp, E., Teperek, M., Turkyilmaz-van der Velden, Y., & Versteeg, A. (2019). Policy needs to go hand in hand with practice: The learning and listening approach to data management. *Data Science Journal*, 18(1), 1–11. <https://doi.org/10.5334/dsj-2019-045>

Digital Science; Fane, B., Ayris, P., Hahnel, M., Hrynaszkiewicz, I., Baynes, G., & Farrell, E. (2019). The State of Open Data Report 2019 (version 2). <https://doi.org/10.6084/m9.figshare.9980783.v2>

Erway, R. (2013). Starting the conversation: University-wide research data management policy. *OCLC Research*, Dublin, Ohio. <http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf>

Felder, L. (2011). *Writing for the web: Creating compelling web content using words, pictures, and sound*. Que Publishing.

Green, A., Macdonald, S., & Rice, R. (2009, May). Policy-making for research data in repositories: A guide. EDINA. <http://www.disc-uk.org/docs/guide.pdf>

Higman, R., & Pinfield, S. (2015). Research data management and openness the role of data sharing in developing institutional policies and practices. *Program Electronic Library and Information Systems*, 49(4), 364–381. <https://doi.org/10.1108/PROG-01-2015-0005>

Jones, S., Ross, S., & Ruusalepp, R. (2008, September 29-30). The Data Audit Framework: A toolkit to identify research assets and improve data management in research led institutions. 5th International iPRES Conference (“Joined Up and Working: Tools and Methods for Digital Preservation”), London, England. pp. 213–219. <https://eprints.gla.ac.uk/6240/1/6240.pdf>

Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLOS One*, 12(3). <https://doi.org/10.1371/journal.pone.0173987>

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L’Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolva, D. V., Stockhause, M., & Khodiyar, V. (2020). The TRUST Principles for digital repositories. *Scientific Data*, 7(1), 1–5. <https://doi.org/10.1038/s41597-020-0486-7>

Lippencott, S. (2017). Library as publisher: New models of scholarly communication for a new era. *Against the Grain Press*. <http://dx.doi.org/10.3998/mpub.9944345>

McCready, K., & Molls, E. (2018). Developing a business plan for a library publishing program. *Publications*, 6(4), 42. <https://doi.org/10.3390/publications6040042>

Patterson, L., Bothma, T. J. D., & Deventer, M. J. (2018, January 1). From planning to practice: An action plan for the implementation of research data management services in resource-constrained institutions. *Library and Information Association of South Africa (LIASA)*. https://www.openaire.eu/search/publication?articleId=od_1575::c3f38b68013b129e38b58fe160809bea

Peoples, B. K., Midway, S. R., Sackett, D., Lynch, A., & Cooney, P. B. (2016). Twitter predicts citation rates of ecological research. *PLoS ONE* 11(11), e0166570. <https://doi.org/10.1371/journal.pone.0166570>

Piwovar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, 1, e175. <https://doi.org/10.7717/peerj.175>

- Riddle, K. (2015). Creating policies for library publishing in an institutional repository. *OCLC Systems & Services*, 31(2), 59–68. <https://doi.org/10.1108/OCLC-02-2014-0007>
- Schaeffer, P. (2011, October 5). Why does Dryad use CC0? Dryad News and Views, Dryad [blog post]. <https://blog.datadryad.org/2011/10/05/why-does-dryad-use-cc0/>.
- Shearer, K. (2015, April 7). Comprehensive brief on research data management policies. <https://portagenetwork.ca/wp-content/uploads/2016/03/Comprehensive-Brief-on-Research-Data-Management-Policies-2015.pdf>
- SPARC Europe (2017, February 28). The open data citation advantage [Briefing Paper]. <https://sparceurope.org/open-data-citation-advantage/>.
- Tenopir, C., Talja, S., Horstmann, W., Late, E., Hughes, D., Pollock, D., Schmidt, B., Baird, L., Sandusky, R., & Allard, S. (2017). Research data services in European academic research libraries. *Liber Quarterly: The Journal of the Association of European Research Libraries*, 27(1), 23–44. <https://doi.org/10.18352/lq.10180>
- Van Zeeland, H., & Ringersma, J. (2017). The development of a research data policy at Wageningen University & Research: Best practices as a framework. *Liber Quarterly: The Journal of the Association of European Research Libraries*, 27(1), 153–170. <https://doi.org/10.18352/lq.10215>
- Verhaar, P., Schoots, F., Sesink, L., & Frederiks, F. (2017). Fostering effective data management practices at Leiden university. *Liber Quarterly: The Journal of European Research Libraries*, 27(1), 1–22. <https://doi.org/10.18352/lq.10185>
- Yu, H. H. (2017). The role of academic libraries in research data service (RDS) provision: Opportunities and challenges. *The Electronic Library*, 35(4), 783–797. <https://doi.org/10.1108/EL-10-2016-0233>

APPENDIX A

List of Policies and Guidance

University or Organization Name	Link to Policy or Guidance
Cornell University	https://guides.library.cornell.edu/ecommons/datapolicy
Harvard Dataverse	https://support.dataverse.harvard.edu/policies
MIT	https://libguides.mit.edu/c.php?g=176372&p=1158986
Oregon Health & Science University	https://scholararchive.ohsu.edu/about?locale=en
Purdue University	https://purr.purdue.edu/legal/terms
Rutgers	https://www.libraries.rutgers.edu/services-for-researchers/data-services/nb-data-management-services
Syracuse University Qualitative Data Repository	https://qdr.syr.edu/policies
University of Arizona	https://data.library.arizona.edu/data-management/best-practices/data-sharing-archiving
University of Minnesota	https://conservancy.umn.edu/pages/drum/policies/
University of Nebraska-Lincoln	https://dataregistry.unl.edu/researchers.html#Preservation
Washington University in St. Louis	https://openscholarship.wustl.edu/data/policies.html

APPENDIX B

Comparison of Initial Section Titles to the Posted Version's Section Titles

The table below details our initial headings, how these headings were modified or changed, and the final headings to our first posted policy (posted in 2019). Our current policy is available at <https://scholarworks.umass.edu/data/policies.html>.

Initial Heading— Gathered From Policy Review	Review of Policy and Comments on Heading and Section	Posted Policy Headings and Sections
“Table of Contents”	Unchanged	“Table of Contents”
“General Statement”	Unchanged	“General Statement”
“Data Collection Policy”	Modified for clarity	“Data We Collect (Data Collection Policy)”

(Table continues on following page)

Initial Heading— Gathered From Policy Review	Review of Policy and Comments on Heading and Section	Posted Policy Headings and Sections
“End-User Access Policy”	Modified for clarity	“How People Can Use Your Material (End-User Access Policy)”
“Deposit License”	Modified for precision	“What You Need To Agree To In Order For Us To Host and Share Your Work (Deposit License)”
“Submission Agreement”	Removed from policy; this is used elsewhere in the data deposit workflow	
“Terms of Use”	Removed from final policy—out of scope	
“Takedown Policy”	Removed from final policy—true takedown policy is rigorous and requires a great deal of knowledge of copyright law (which is beyond the scope of our work)	
“Boilerplate language for grants”	Modified for clarity	“Language for Use in the Grant- Writing Process”
“Registering Data”	Modified for precision	“Registering Your Data in ScholarWorks”
“Policies cited in the creation of this document”	Changed the word “cited” to “used,” because our policy was based off other policies; the phrase “cited” implies incorporation of other policies (and policies at other organizations) into our own	“Policies used in the creation of this document”

APPENDIX C
Submission Form Fields

Field	Description	Required (Y/N)
Title	The dataset’s title	Y
Authors	Author lookup via email or manual entry	Y
Publication Date	Date of publication; only the year is required	Y
Keywords	Keywords that help describe the dataset; helpful in improving retrieval by search engines; separated by comma	Y

(Table continues on following page)

Field	Description	Required (Y/N)
Disciplines	Discipline(s) under which this work falls	N
Description	A brief description of the data; focus is on helpful details about the data that may help improve reuse	N
Digital Object Identifier (DOI)*	The administrators mint Digital Object Identifiers on behalf of depositors; this step is not visible to depositors and occurs after the data are initially queued for deposit in the system	N
Grant/Award Number and Agency	Grant or award number and the funding agency	N
Primary Publication Related to this Data	Digital Object Identifier or web page address of one article the author would like to associate with this dataset	N
Additional Related Content	Open text that accepts HTML to link to related content; could include links to code, other articles, websites, or other content	N
Document Type	Dropdown box, defaults to “Data”; alternative option is “None”	N
Rights	Open text to document relevant copyright or usage rights	N
Creative Commons License	Dropdown box, defaults to “None”; includes two Creative Commons license options: Creative Commons Public Domain Dedication 1.0 (CC0) and Creative Commons Attribution 4.0 (CC-BY)	N
Recommended Citation	Open text box; Digital Commons automatically generates a suggested citation ; use of this feature will override the default citation	N
Upload Data File	Prompt for how user will upload a file. Radio button to select one option of three: <ul style="list-style-type: none"> • Upload file from your computer • Import file from remote site • Link out to file on remote site 	Y
Cover Image	Prompt to select a cover image for the dataset; if the user does not select a custom image, the default image for a dataset in the data repository is applied	N
README File and Additional Files	Checkbox—If user has not included a readme file with their data already, a check in this box will prompt the system to allow for additional files to be uploaded	N
Embargo Period*	If requested, we can apply an embargo to the data; however, there is no nuance to our embargo parameters, so we are unable to allow for specific access to a certain individual or group; this is an “all-or-none” condition	N

* Denotes that this field is only for administrator use and is hidden from public view.

APPENDIX D

Screen Capture of Data Deposit page—User's View

ScholarWorks@UMass Amherst (<https://scholarworks.umass.edu>)

[DATA AND DATASETS \(HTTPS://SCHOLARWORKS.UMASS.EDU/DATA\)](https://scholarworks.umass.edu/data)

REQUIRED Title

Please enter the title of your data set or collection

Enter title:

REQUIRED Authors

Search For An Author Using: Last Name, First Name, Email, or Institution

1 (tpatwood@umass.edu) Thea P Atwood, University of Massachusetts Amherst +

REQUIRED Publication Date

Only the year is required but all other fields are optional.

Season: Month: Day: Year:

REQUIRED Keywords

Please enter keywords or phrases, separated by commas. Keywords and descriptors help identify and classify your dataset. This helps enhance how it is retrieved by search engines. For example, you might list geographic locations or scientific names not mentioned in the title or your description.

Enter keywords:

Disciplines

Available:

- Architecture
- Arts and Humanities
- Business
- Education
- Engineering
- Law
- Life Sciences
- Medicine and Health Sciences
- Physical Sciences and Mathematics
- Social and Behavioral Sciences

Selected:

Figure D1. a) User view of the Data & Datasets repository submission form, part 1 of 3. b) User view of the Data & Datasets repository submission form, part 2 of 3. c) User view of the Data & Datasets repository submission form, part 3 of 3.

Description

Please include a brief description of your data. Helpful details may include the purpose of the study; information about the study design and methodology; how the data was generated; key dates associated with the data; information about its spatial coverage

Paste or type your description:

B I | | | | HTML

| |

Grant/Award Number and Agency

If your data is associated with a grant or award, please enter the grant/award number and the name of the funding agency.

B I | | | | HTML

| |

Primary Publication Related to this Data

Please enter the URL or DOI of one publication you would like associated with this dataset

Additional Related Content

Enter links to content directly related to this data such as articles or other data for this project. Use [HTML hyperlinks](#) (http://www.w3schools.com/tags/tag_a.asp), or the Link button in the controls. Please enter one link per line.

B I | | | | HTML

| |

Document Type

Type of submission ▼

Figure D1. *Continued*

Rights

Enter relevant copyright and/or usage rights information for this submission

Creative Commons License

Recommended Citation

A recommended citation is automatically generated for all submissions. To override the default suggested citation, please enter citation information below

Upload Data file

Please upload a file from your computer or enter the URL that links to your data set at an external archive or repository. Note that you can zip together many files into one data package

- Upload file from your computer
- Import file from remote site
- Link out to file on remote site

Cover Image

Select an image that represents your data. It will be displayed as a thumbnail on the website. If you do not select an image, a default image will be provided. Selecting the "1st page of PDF" option will prompt the system to provide a default image.

- Use 1st page of PDF
- Upload a cover image

README file and Additional Files

Please check this box to add your README file if it is not already in your data package.

This may take a while. **Please only click once.**

Figure D1. Continued