Engaging Researchers in Data Dialogues: Designing Collaborative Programming to Promote Research Data Sharing

Moira Downey
Duke University

Et al.

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/jeslib

Part of the Scholarly Communication Commons, and the Scholarly Publishing Commons

Repository Citation
https://escholarship.umassmed.edu/jeslib/vol10/iss2/3

Creative Commons License
This work is licensed under a Creative Commons Attribution 4.0 License.
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMassChan.
For more information, please contact Lisa.Palmer@umassmed.edu.
Abstract

A range of regulatory pressures emanating from funding agencies and scholarly journals increasingly encourage researchers to engage in formal data sharing practices. As academic libraries continue to refine their role in supporting researchers in this data sharing space, one particular challenge has been finding new ways to meaningfully engage with campus researchers. Libraries help shape norms and encourage data sharing through education and training, and there has been significant growth in the services these institutions are able to provide and the ways in which library staff are able to collaborate and communicate with researchers. Evidence also suggests that within disciplines, normative pressures and expectations around professional conduct have a significant impact on data sharing behaviors (Kim and Adler 2015; Sigit Sayogo and Pardo 2013; Zenk-Moltgen et al. 2018). Duke University Libraries' Research Data Management program has recently centered part of its outreach strategy on leveraging peer networks and social modeling to encourage and normalize robust data sharing practices among campus researchers. The program has hosted two panel discussions on issues related to data management—specifically, data sharing and research reproducibility. This paper reflects on some lessons learned from these outreach efforts and outlines next steps.

Correspondence: Moira Downey: moira.downey@duke.edu
Received: September 10, 2020 Accepted: December 18, 2020 Published: March 1, 2021
Copyright: © 2021 Downey et al. This is an open access article licensed under the terms of the Creative Commons Attribution License.
Disclosures: The authors report no conflict of interest.
Introduction

The fourth paradigm of scientific discovery relies upon the generation, collection, and management of digital data (Hey et al. 2009). Sharing these digital data not only advances the research endeavor but also supports reproducibility and integrity. A current shift in academia prioritizes data accessibility and reuse and is expressed through policies from funding agencies (NSF 2011, NIH 2021) and journals (PLOS 2014) as well as through normative shifts within individual disciplines and data communities (Cooper and Springer 2019). However, realizing the benefits of the FAIR (i.e., Findable, Accessible, Interoperable, and Reusable) Guiding Principles for scientific data (Wilkinson et al. 2016) requires cyberinfrastructure resources and data curation expertise to effectively manage data from collection to publication. In response to this growing need, academic libraries have been developing research data management programs (Fearon et al. 2013). Research data management (RDM) can be broadly defined as caring for data throughout a research project to ensure it can be understood, accessed, reused, and preserved for the long term. By making data more broadly available, RDM democratizes both access to knowledge and the practice of data sharing beyond the idiosyncrasies of specific subfields.

Foundational to supporting RDM within an academic library setting is establishing a broad understanding of researchers’ data management and sharing practices, perceptions, and motivations. Over the past two decades, library and information science scholars have explored a wide range of research questions using quantitative and qualitative methods. Multinational surveys have provided a view of how researchers' self-reported practices and perceptions surrounding data sharing and reuse have changed over time (Tenopir 2011; Tenopir 2015; Tenopir 2020). Explorations of the influence of institutional and individual factors on data sharing behaviors point to the complexity of motivating data sharing and the impact of social-normative pressures (Kim & Stanton, 2015). Studies with a disciplinary focus provide reference points for understanding disciplinary norms, differences across disciplines, and the ethics of sharing certain types of data (Cragin et al. 2010; Faniel, Kriesberg, and Yakel 2015; Mozersky et al. 2020; Piwowar 2011; Whyte and Pryor 2011). The current data management and sharing literature exploring the behaviors and motivations of researchers provides a robust view of the data sharing landscape at a macro-level.

Turning within an institution, studies of research communities facilitate understanding local researchers’ needs in order to prioritize services (Akers and Doty 2013; Weller and Monroe-Gulick 2014). Likewise, the expansive literature exploring academic library RDM programs provides a basis for exploring what services are commonly provided, assessment strategies, and roles and responsibilities (Coates et al. 2018; Cox et al. 2017; Hudson-Vitale et al. 2017; Tenopir et al. 2014). Today an increasing role for libraries within institutions is to advocate for normative change and to engage local communities in conversations. Situating these conversations in academic libraries logically extends their changing profile. Having long served as a "crossroads for intellectual activity" on campus,
recent decades have seen academic libraries expand their focus beyond acquiring and providing access to research materials—namely books and print-based journals—to serving as a site of support for activities spanning the research lifecycle (Council on Library & Information Resources 2008). Libraries have endeavored to reshape their campus image into that of an active partner in research and to restructure their prevailing service models to accommodate the full scope of research activity (Dempsey and Malpas 2018; Vaughan et al. 2013; Wynne et al. 2016). Some have even reconfigured physical spaces to better reflect this shift, opening up areas traditionally reserved for print collections to better collocate research services (Latimer 2011).

This growing and changing service portfolio, in conjunction with academic libraries' existence at the intersection of disciplines, lends them particular weight in shifting research norms. A challenge remains, however, in engaging the still sizable portion of campus researchers who continue to view the primary role of the libraries to be the purchase of research materials (Schonfeld and Wulfson 2014). Strategies for engagement and communication with user communities can take various forms and include education and training, presentations, attending external events, online communication, contributing to institutional initiatives and groups, and sponsoring programming and events. However, as Latham (2017) notes “while outreach/promotion and collaboration are recognized as integral to pairing communities of users with services of value to their research, when it comes to RDM, outreach is seldom afforded primacy” (264). And fundamental to effective outreach is formulating a communication strategy that engages the research community.

There is an expansive theoretical toolkit that may be used to inform the development of a specific communication strategy, particularly when a desired outcome centers on reshaping behaviors. As established by Fisher (1984), narrative paradigm theory purports that the primary mode of human communication is through storytelling and that stories are "meant to give order to human experience and to induce others to dwell in them to establish ways of living in common" (6). The technique of narrative in effecting behavioral change has been well-studied, particularly within the field of healthcare and health outcomes (Hinyard and Kreuter 2007). Storytelling is an effective sense-making tool, and individuals often narrativize their own lives in order to better understand them (Rindfleish, Sheridan, and Kjeldal 2009). In allowing individuals to share their personal contexts, storytelling creates "conditions in which people's co-constructed worlds of meaning are spontaneously revised in interaction" (Shaw 1997, 179), and as such, stories can also serve as "especially viable instruments for social negotiation" (Bruner 1990, 97). Given the efficacy of stories for shaping behaviors, coupled with evidence of the importance of social-normative pressures in motivating data sharing (Kim and Stanton 2015), we found a communication strategy rooted in narrative paradigm theory to be an appealing one. This model builds upon other RDM efforts where libraries facilitate communication and form relationships with researchers (Murray and Carson 2018).
Research Data Management at Duke

In addition to the growing regulatory pressures posed by funding agencies and journals and the general needs presented by the changing nature and scope of scholarly research, Duke University has faced some recent context-specific challenges. Accusations of research misconduct over the course of the last decade have brought the University under increased scrutiny from a number of federal funding agencies (Luzum 2019). In an effort to re-center the importance of research integrity on campus, the University has subsequently embarked on a series of initiatives to better ensure the responsible conduct of scientific research, including establishing an Office of Research and appointing a vice dean and associate vice provost for scientific integrity. This position oversees the Advancing Scientific Integrity, Services and Training (ASIST) office as well as supervising the assessment, investigation, and reporting of research misconduct. Recognizing the need to support faculty in both the conduct of responsible research and the stewardship of an increasing amount and array of digital research output, in 2015 a faculty working group that included a number of campus faculty, IT administrators, and librarians, recommended the creation of a research data management and curation program to be overseen by the Libraries. Four new FTEs were onboarded in 2017, at which time the new staff began work to create a suite of data management and curation services, including policies and procedures, while simultaneously rethinking the software infrastructure required.

Throughout early 2017, staff established a pre-publication curation workflow for ensuring the quality of submitted datasets and began work with library software developers to create a new platform dedicated to the publication and preservation of research data. In crafting a service profile and building a new repository application to support data publication, the curation team relied heavily on best practices as outlined by the literature and put forward by organizations such as the Data Curation Network (Lafferty-Hess, et al. 2020). Duke’s Research Data Management program as presently constituted provides three major areas of support: building knowledge and skills through education and outreach; meeting the research data management needs of scholars throughout the research lifecycle and offering assistance with data workflow questions; and a data curation, publication, and preservation program built around a locally managed data repository.

Research Data Management Outreach

Once the key elements of Duke’s RDM program were established, additional effort was dedicated to institutional outreach. The Duke RDM program has taken a multi-faceted approach to outreach, advocacy, and communication with our research community. We have partnered with various research units to target specific groups such as working with the Graduate School to offer data management Responsible Conduct of Research (RCR) workshops. These partnerships proved a useful strategy to train graduate students while simultaneously raising awareness of available services, as well as contributing to
events hosted by Research Computing and the Office of Scientific Integrity. Likewise, we have hosted events for groups on campus, including IT, research staff, and grant managers. We have also presented during faculty departmental meetings.

Another strategy employed within the libraries has been hosting panel events where researchers share their experiences on a particular data management topic. While the examples of outreach described above have been productive, they are largely one-way communications where program staff educate or “pitch” our research community on available services. By contrast, the library-based panels engage researchers as active participants and contributors and focus on peer-to-peer learning and social modeling. Duke University Libraries has hosted two panel events, one on data sharing and one on reproducible research.

**Data Management Panel Discussions**

In April 2019, we hosted our first panel, featuring faculty from three academic disciplines: Evolutionary Anthropology, Civil and Environmental Engineering, and Chemistry. We asked them to discuss their personal experiences with data sharing, why this practice has been important in their careers, and their general perceptions on how to further encourage data sharing. All three had varying reasons for sharing, including ensuring data access through redundancy, maintaining consistency in lab workflows in the face of frequent turn-overs in student staffing, and facilitating collaboration among geographically dispersed colleagues. This panel was also included for one credit hour under the Office of Scientific Integrity Responsible Conduct of Research (RCR) program for faculty and staff.

We followed up our data sharing panel with another discussion focused on reproducibility in practice in Fall 2019. Taking a 2016 article from Nature about the reproducibility crisis (Baker 2016) as a point of departure, panelists discussed some of the key challenges to making research reproducible, what tools and methods they have used to do so, and how to teach early-career researchers about the importance of reproducibility. Panelists were intentionally drawn from across multiple disciplines including Biostatistics and Bioinformatics, Marine Science and Conservation, Information Science, and Statistical Science. For others interested in hosting similar events the details of our process are described below.

The general process for developing both panels included first identifying relevant panelists and sending invites via email. For the first panel we drew on faculty with whom we had existing relationships and knew were engaged in data sharing practices, either locally through our institutional repository or through disciplinary repositories. For the second panel we turned to another set of researchers with whom we had established relationships, and, in addition, successfully invited two panelists who we knew were interested in reproducibility, but with whom we had no prior contact. In preparation, we provided panelists with potential questions and topics. For the data sharing panel, researchers were given some time to
present their perspective and experiences; by contrast, the reproducibility panel did not include structured presentations. An RDM consultant moderated both panels by presenting the prepared questions, but a significant amount of time was reserved for open questions and discussion with the audience. Marketing of the sessions was primarily conducted through the Libraries’ Center for Data and Visualization Sciences listserv, social media outlets, digital signage on campus, and by providing a free lunch.

We have provided two narrative case studies illustrating the kind of information and peer-to-peer sharing that took place during the data management panels as supplemental material to this article (see Appendix). A significant aspect of running a successful panel relies upon recruiting willing researchers with nuanced and real-world experience with the topic at hand. Therefore, the motivations of panelists to contribute to such panels is highly relevant when approaching future participants. From our case studies, we have highlighted two panelists’ views of the value of participating in these types of events below.

**Dr. Charbonneau:** Changing research practices can be difficult. Why would one invest energy in altering something that does not immediately increase research funding? The truth is that data management practices are steadily shifting and there’s a real risk of losing competitiveness. Being at the forefront of that evolution allows us to set the standards that others must follow, rather than the opposite. Explaining such non-quantifiable benefits, however, takes time, and might not necessarily be a net gain for every single researcher. Guiding colleagues as they take their first step toward deposition is my way of giving back for the peace of mind depositing research products has brought to my faculty career.

**Dr. Zoss:** Just as reproducibility itself inspires reflection on the research process, sharing my thoughts on reproducibility as part of a panel inspired reflection on my own reproducibility process. It gives me an opportunity to renew my commitment to reproducibility and to learn about new tools and techniques from others. It has helped me feel more a part of a community, when research can sometimes encourage isolation and intense focus. Just as reproducibility itself can alleviate the pressure to do everything at once (i.e., before you forget), taking time to pause and share as a community helps build a culture around reproducibility and encourages people to re-examine and update aspects of their work that may not be as visible or as valued by colleagues.

**Lessons Learned and Next Steps**

**Assessment**

Assessment of these communication efforts at DUL has been lightly reverse-engineered; we did not embark on this communication strategy with a specific assessment instrument in mind. Despite the lack of a formal framework for
evaluating the success of our panels, we were able to capture some registration data that provides some insight into interest in and engagement with this kind of outreach. Both panels met the maximum number of registrants (40), and with waitlists of 20 and 18 enrollees, respectively, these events represented two of our more popular sessions. Registrants skewed heavily toward staff affiliated with the medical side of the University or the Duke University Health System, but faculty, graduate students, and postdoctoral researchers (postdocs) from a wide variety of disciplines were also represented, giving us a modest cross-section of campus (see Fig. 1). Clustering from the medical side of campus may be an artifact of heightened regulatory scrutiny from external funders focused on those disciplines that is specific to Duke’s context. No undergraduates are recorded among the attendees, which may be attributable to the panels’ inclusion in the RCR credit program that is oriented toward faculty, staff, and graduate students. Although we did not capture a more granular formal classification of the staff registrants, several library staff members attended along with research services staff from across the University. Finally, we have noticed a trend wherein deposits seem to cluster around lab or research group members or around departmental affiliation, which may further suggest a networked, word-of-mouth phenomenon. As we move forward with this approach, we intend to coordinate with DUL’s Assessment and User Experience unit to construct a formal assessment framework and situate this work within our broader RDM program goals and ongoing evaluation.

<table>
<thead>
<tr>
<th>Disciplinary grouping or affiliation</th>
<th>Faculty</th>
<th>Graduate students</th>
<th>Postdocs</th>
<th>Staff</th>
<th>Other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEM</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>9</td>
<td>1</td>
<td>22</td>
</tr>
<tr>
<td>Humanities &amp; Social sciences</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>Medical &amp; Public health</td>
<td>11</td>
<td>3</td>
<td>1</td>
<td>34</td>
<td>2</td>
<td>51</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>1</td>
<td>29</td>
<td>1</td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>12</td>
<td>6</td>
<td>75</td>
<td>4</td>
<td>118</td>
</tr>
</tbody>
</table>

**Figure 1:** Panel registrants by academic status and disciplinary grouping
Outreach and Advocacy

As academic libraries’ positions on campus shift from a place where researchers acquire materials to a place where researchers gain support for various aspects of data intensive scholarship, outreach to user communities also must shift. Libraries are no longer just service provision units but partners enabling more reproducible, responsible, and impactful research. While this shift has been taking place over decades, the perception of libraries’ role on campus is still in flux (Bryant, Dortmund, and Lavoie 2020). Relevancy relies upon a commitment to engaging in a meaningful way with the campus community and positioning the library as an active member. The Duke RDM panel discussions allowed us to raise awareness of available services, such as the Duke Research Data Repository, without directly “pitching” said services. By providing the space for our research community to engage in dialogues on timely RDM topics, researchers themselves placed library services within the context of their own experiences. We can build service portfolios and platforms for our communities, but collaborations with our research communities are often the most fruitful strategies for effective outreach and advocacy. For example, as a faculty champion, one of this paper’s authors has not only participated in data management panels and shared his personal story through this article, but also supported platform development through user feedback, encouraged library presentations at departmental meetings, and contributed to various online outreach initiatives. These types of relationships with faculty and other campus groups are foundational for ongoing outreach.

Peer-to-peer Engagement

Given the role of social-motivational and sociocultural influences on knowledge acquisition (O’Donnell and King 1999), and particularly the impact of social normative pressures on positively impacting data sharing (Kim and Stanton 2015), it is not surprising that we found it effective to have researchers talk to each other versus library staff talking at them. Framing the panels around how data sharing and reproducible research practices were established within their own labs or research projects elicited candid reflection and open sharing of decision points, techniques, and challenges. Likewise, the various voices and disciplines represented on the panels gave space to explore a wide range of ideas and approaches related to these complex topics. By loosely rooting our communication strategy in narrative paradigm theory, we allowed researchers to present their own individualized stories as a sense-making strategy for how data sharing and reproducibility are actually implemented within specific research contexts.

Engaging the Humanities

Perhaps unsurprisingly, but nonetheless notably, as measured by deposits and consultations, campus-wide engagement remains strongest among the physical sciences, engineering, and health sciences. While a number of infrastructural and behavioral obstacles remain regarding sharing humanistic data, the growth in digital scholarship in the humanities provides new opportunities for engaging
humanistic scholars in conversations regarding data sharing (Poole 2017; Almas 2017). Within the context of DUL’s RDM programming, we had planned to attempt to address this discrepancy by pivoting to a more targeted session this past spring with a session entitled “Beyond the manuscript: Sharing humanistic ‘data’ in the digital age,” featuring researchers from Art, Art History & Visual Studies, History & Law, and English. This session was being planned in collaboration with the Libraries’ Digital Scholarship Services Department and ScholarWorks. While we unfortunately had to postpone the session in the face of campus shutdowns related to the COVID-19 global pandemic, outreach to the humanities remains a top priority for us moving forward.

Conclusion

As academic libraries continue to define and refine their role in the research data management and curation space the challenge of engaging scholars and their broader communities remains forefront. Outreach is not easy, but it is an essential piece of the work. Libraries can enlist researchers themselves in efforts to meet this challenge. Encouraging them to narrativize their own experiences helps both to concretize often abstract concepts by situating them within actual research workflows, and to normalize data sharing behaviors. Through fostering and hosting these conversations, we position the library as active partners in these dialogues and strengthen relationships across campus.

It should be acknowledged, however, that there are hurdles to taking this approach. Sustainability of these efforts requires resources, including a non-trivial amount of time from staff who have an array of other professional responsibilities. Likewise, relying on existing RDM staff connections may limit our ability to recruit new (and willing) participants and highlights the value of collaborating with others in the library to harness broader networks. Staying relevant and reaching new audiences, such as humanities scholars, also relies upon understanding the particular issues and stories they may wish to tell or engage. As we look toward the future, we plan to stay nimble in our approach and formalize our evaluation, while also looking for opportunities to position these conversations within the context of changes in the research landscape, such as the recently released NIH Data Management and Sharing Policy.

Acknowledgements

This paper is an extension of the Poster “Reaching Across Campus: Designing Cross-Disciplinary Programming to Promote Research Data Sharing” presented at the 2020 Research Data Access and Preservation (RDAP) Conference. Patrick Charbonneau was supported by a National Science Foundation Grant No. DMR-1749374. We would also like to thank our colleagues, Jennifer Darragh and Shadae Gatlin, for their feedback on this paper.
Supplemental Content

Appendix
An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2021.1193 under “Additional Files”.

References


Engaging Researchers in Data Dialogues


Mozersky, Jessica, Heidi Walsh, Meredith Parsons, Tristan McIntosh, Kari Baldwin, and James M. DuBois. 2020. "Are We Ready to Share Qualitative Research Data? Knowledge and Preparedness among Qualitative Researchers, IRB Members, and Data Repository Curators." IASSIST Quarterly 43 (4). https://doi.org/10.29173/iq952


Engaging Researchers in Data Dialogues


