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

Authors: Moira Downey (Duke University Libraries), Sophia Lafferty-Hess (Duke University Libraries), Patrick Charbonneau (Duke University), Angela Zoss (Duke University Libraries)

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 this outreach effort and outlines next steps.

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. We would also like to thank our colleagues, Jennifer Darragh and Shadae Gatlin, for their feedback on this paper.

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 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 (please 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: To now, any 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, however, 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 and graduate or postdoctoral students 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.

Disciplinary grouping or affiliation	**Career status**					
	Faculty	**Graduate student**	**Postdoctoral student**	**Staff**	**Other**	**Total**
STEM	6	4	2	9	1	22
Humanities & Social sciences	3	5	2	3		13
Medical & Public health	11	3	1	34	2	51
Other	1		1	29	1	32
Total	**21**	**12**	**6**	**75**	**4**	**118**

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, 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.

Appendix

Patrick Charbonneau: Data Sharing Panel

Patrick Charbonneau conducted his undergraduate study at McGill, in Montréal, obtained his Ph.D. in chemical physics from Harvard University in 2006 and then was a Marie-Curie Postdoctoral Fellow at Amolf, in Amsterdam, before joining Duke in 2008, where he is now associate professor of chemistry and physics.

Practices: How have you integrated data sharing and archiving into your workflows or research practices?

Dr. Charbonneau: When I started at Duke, my archival strategy consisted of waiting until researchers in my group were about to leave before having them save and document their data to a backed up departmental drive. Because that approach was a marked improvement over how I managed research material during my own graduate and postdoc years, I was proud to disclose it in early NSF data management plans (Pasek 2017). The passing years, however, revealed its weaknesses. Out of the blue, I would receive requests for data whose creator had long since left Duke. Handling these requests was stressful because our retroactive and last-minute metadata practices were far from comprehensive. Fortunately, because former students had remained in touch we muddled through, but that was more luck than forethought. Before luck ran out, I sought to implement a model in which (i) former affiliates (or myself) would need not be involved, and (ii) archiving could be handled on a per-publication basis, when everything is fresh. Since 2016, with the help of Duke Libraries, my group has thus developed a protocol by which each publication could cite a DOI-labeled repository containing code, data, and archival-quality metadata. The key to success is having this material at hand by the time the publication is accepted, then promptly passing the data on to the Duke Research Data Repository curators and inserting the repository info in the article at the proofing stage. To make sure this choreography goes smoothly it is now an expectation in my group that in parallel to manuscript preparation, raw data and codes are also getting

ready for public release. Interestingly, this routine also provides an additional validation route of the research data before publication.

Perceptions: What are the key challenges or opportunities that everyone in research in your field (or any field) should know about related to data sharing/reproducibility? What do you feel are the most effective ways to engage around the topic of data sharing/reproducibility?

Dr. Charbonneau: Shifting from the old to the new model of data archiving mostly involved overcoming a logistical (or cultural) barrier. There was nothing fundamentally new in our values or intents. Setting a protocol in place, with curatorial oversight, merely kept our practices in line with these ideals. The challenge is therefore to do it for the first time. Afterwards, for students and postdocs this practice becomes the *proper* way to publish research findings.

The key practical question is deciding what to include. All raw data or only (partially) processed data? Should one be concerned as to whether this data might or might not be reused? There's no right answer to these questions, but setting small goals facilitates getting started, which I feel is the most important consideration. Aiming for the lowest common denominator also broadens disciplinary relevance. Scholars from history to biology to economics and literature publish plots with points, whose coordinates could be deposited. After a few deposition cycles, the scope can grow and specialize, in response to collegial requests or otherwise. It's also possible to go back and extend the content of an initial deposition. Version control makes that completely transparent.

The most effective—non-coercive—way to encourage the use of repositories is emphasizing that depositing research products comes at essentially no cost and offers key benefits. Research productivity does not suffer because depositing research products is interwoven into the manuscript and figure preparation process. It mostly reorders some of the steps most researchers already take. The benefits, however, are substantial. As mentioned above, it provides an additional guardrail for pre-publication data handling, and it reduces the need to handle individual data requests. It can also improve the long-term impact of a research effort. Who knows what use any given particular dataset could have for future

researchers? I at least know that some of my requests for old codes and data have been unfulfilled because of agedness, of either technology or researchers.

Angela Zoss: Reproducibility Panel

Angela Zoss holds a Master's degree in Communication from Cornell University and a Ph.D. in Information Science from Indiana University, where she also completed her undergraduate study. She has worked at Duke University since 2012, first as a Data Visualization Coordinator and now as an Assessment and Data Visualization Analyst in the Duke University Libraries.

Practices: How have you integrated reproducibility into your workflows or research practices?

Dr. Zoss: From 2012 to 2018, I was working a full-time job while trying to conduct and complete my doctoral dissertation research. I found it difficult to reserve significant and consistent blocks of time for my research, and my committee members and I would struggle to remember what previous decisions had been made about data collection and analysis. As I had been learning and teaching R during my "day job," I decided to use my dissertation research as an opportunity to improve my proficiency with that tool.

I started by using R to analyze the results of a simple pilot survey. When I realized the many benefits of a fully reproducible analysis workflow, I sought out a framework that would help me organize the various data and script files I was generating. Someone recommended the TIER Protocol[1] to me, and I found it to be a wonderful fit for the kind of analysis work I was doing. Combining R projects and the TIER Protocol, I was able to generate a series of analysis workflow steps that took me all the way from the preparation of components included in my main experiment through the data cleaning of participant responses to the final data modeling and visualizations for my dissertation.

[1] https://www.projecttier.org/tier-protocol/

After completing my Ph.D., I have continued to use reproducible workflows in my full-time position as an assessment and data visualization analyst. Any project requiring data cleaning, blending, visualization, and publication is made much simpler by standard reproducibility practices. I have continued to use R for many projects, but I also prioritize other types of reproducible workflows like command-line scripts or tools with exportable steps when R is not appropriate. These practices benefit not only my project partners and me, but also a much broader community. One large project that required gathering and blending data from a variety of sources was recently presented at a national conference. After the presentation, I was immediately approached by someone who thanked me for sharing the files openly on GitHub.

Perceptions: What are the key challenges or opportunities that everyone in research in your field (or any field) should know about reproducibility? What do you feel are the most effective ways to engage with the topic of reproducibility?

Dr. Zoss: One challenge I see for reproducibility is that exposure to these practices was not centralized in my graduate program, which meant that my understanding of best practices was piecemeal. I was training in a social science discipline where formal statistical analysis and programming for research projects were not widespread practices. I hope in the intervening years these concepts have begun diffusing into all disciplines, but as a new graduate student, finding advice on reproducibility best practices involved haphazard searching and filtering out materials designed either for much more technical work, such as formal software development projects, or for STEM/health fields.

The core principles of reproducibility that I think apply to all fields are as follows: reproducible practices improve research quality, even if the research is never reproduced; reproducibility requires transparency throughout the entire research process; and reproducibility is better served by simple, sustainable documentation than by reliance on complex and ever-changing software. Finding ways to engage those principles, regardless of your discipline or the tools you use for your research, leads to a less stressful, more productive, and higher quality research experience. I highly recommend using a small

project early in your research career to explore what reproducibility means for your work, and then building on that with ideas that come from your own experience as well as that of others in similar fields.

Acknowledgments

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.

References

Akers, Katherine G., and Jennifer Doty. 2013. "Disciplinary differences in faculty research data management practices and perspectives." *International Journal of Digital Curation* 8 (2): 5–26. https://doi.org/10.2218/ijdc.v8i2.263.

Almas, Bridget. 2017. "Perseids: Experimenting with Infrastructure for Creating and Sharing Research Data in the Digital Humanities." *Data Science Journal* 16: 19. http://doi.org/10.5334/dsj-2017-019.

Baker, Monya. 2016. "1,500 scientists lift the lid on reproducibility." *Nature* 533 (7604): 452–454. https://doi.org/10.1038/533452a.

Bryant, Rebecca, Annette Dortmund, and Brian Lavoie. 2020. *Social Interoperability in Research Support: Cross-Campus Partnerships and the University Research Enterprise*. Dublin, OH: OCLC Research. https://doi.org/10.25333/WYRD-N586.

Bruner, Jerome. 1991. "The narrative construction of reality." *Critical Inquiry* 18: 1-21. https://doi.org/10.1086/448619.

Coates, Heather. L., Jake Carlson, Ryan Clement, Margaret Henderson, Lisa R. Johnston, and Yasmeen Shorish. 2018. "How are we Measuring Up? Evaluating Research Data Services in Academic

Libraries." *Journal of Librarianship and Scholarly Communication* 6 (1): eP2226. https://doi.org/10.7710/2162-3309.2226.

Cooper, Danielle, and Rebecca Springer. 2019. "Data Communities: A New Model for Supporting STEM Data Sharing." *Ithaka S+R.* Last Modified 13 May 2019. https://doi.org/10.18665/sr.311396. https://doi.org/10.18665/sr.311396.

Council on Library and Information Resources. 2008. *No brief candle: reconceiving research libraries for the 21st century*. No. 142. Council on Library & Information Resources. http://www.clir.org/wp-content/uploads/sites/6/pub142.pdf.

Cox, Andrew M., Mary Anne Kennan, Liz Lyon, and Stephen Pinfield. 2017. "Developments in research data management in academic libraries: Towards an understanding of research data service maturity." *Journal of the Association for Information Science and Technology* 68 (9): 2182–2200. https://doi.org/10.1002/asi.23781.

Cragin, Melissa. H., Carole L. Palmer, Jacob R. Carlson, and Michael Witt. 2010. "Data sharing, small science and institutional repositories." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368 (1926): 4023–4038. https://doi.org/10.1098/rsta.2010.0165.

Dempsey, Lorcan, and Constance Malpas. 2018. "Academic library futures in a diversified university system." In *Higher education in the era of the fourth industrial revolution,* edited by Nancy W. Gleason, 65-89. Singapore: Palgrave Macmillan. https://doi.org/10.1007/978-981-13-0194-0_4.

Faniel, Ixchel. M., Adam Kriesberg, and Elizabeth Yakel. 2015. "Social scientists' satisfaction with data reuse." *Journal of the Association for Information Science and Technology* 67 (6): 1404-1416. https://doi.org/10.1002/asi.23480.

Fearon, Jr., David, Betsy Gunia, Barbara E. Pralle, Sherry Lake, and Andrew L. Sallans. 2013. *SPEC Kit 334: Research Data Management Services* (July 2013). Association of Research Libraries. https://doi.org/10.29242/spec.334.

Fisher, Walter. R. 1984. "Narration as a human communication paradigm: The case of public moral argument." *Communications Monographs* 51 (1): 1-22. https://doi.org/10.1080/03637758409390180.

Hey, Tony, Stewart Tansley, and Kristin Tolle, eds. 2009. *The fourth paradigm: Data-intensive scientific discovery*. Microsoft Research. https://www.microsoft.com/en-us/research/wp-content/uploads/2009/10/Fourth_Paradigm.pdf.

Hinyard, Leslie J., and Matthew W. Kreuter. 2007. "Using narrative communication as a tool for health behavior change: a conceptual, theoretical, and empirical overview." *Health Education & Behavior* 34 (5): 777-792. https://doi.org/10.1177/1090198106291963.

Hudson-Vitale, Cynthia, Heidi Imker, Lisa R. Johnston, Jake Carlson, Wendy Kozlowski, Robert Olendorf, and Claire Stewart. 2017. *SPEC Kit 354: Data Curation* (May 2017). Association of Research Libraries. https://doi.org/10.29242/spec.354.

Kim, Youngseek, and Melissa Adler. 2015. "Social scientists' data sharing behaviors: investigating the roles of individual motivations, institutional pressures, and data repositories." *International Journal of Information Management* 35 (4): 408-418. https://doi.org/10.1016/j.ijinfomgt.2015.04.007.

Kim, Youngseek, and Jeffrey M. Stanton. 2015. "Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis." *Journal of the Association for Information Science and Technology* 67 (4): 776-799. https://doi.org/10.1002/asi.23424.

Lafferty-Hess, Sophia, Julie Rudder, Moira Downey, Susan Ivey, Jennifer Darragh, and Rebekah Kati. 2020. "Conceptualizing Data Curation Activities Within Two Academic Libraries." *Journal of Librarianship and Scholarly Communication*. 8 (1): eP2347. http://doi.org/10.7710/2162-3309.2347.

Latham, Bethany. 2017. “Research Data Management: Defining Roles, Prioritizing Services, and Enumerating Challenges.” *The Journal of Academic Librarianship* 43 (3): 263–265. https://doi.org/10.1016/j.acalib.2017.04.004.

Latimer, Karen. 2011. “Collections to Connections: Changing Spaces and New Challenges in Academic Library Buildings.” *Library Trends* 60 (1): 112-133. https://doi.org/10.1353/lib.2011.0035.

Luzum, Nathan. 2019. “NIH suspended 7 Duke research grants in 2018 over misconduct allegations and safety concerns.” *The Duke Chronicle*. May 22. https://www.dukechronicle.com/article/2019/05/nih-suspended-duke-university-research-grants-research-misconduct-patient-safety-concerns.

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/iq952Murray, Matthew and A.L. Carson. 2018. “Space for Listening: using a library unConference as an alternative method of communication.” *Journal of EScience Librarianship* 7 (3): e1153. https://doi.org/10.7191/jeslib.2018.1153.

National Institutes of Health. 2021. “Final NIH Policy for Data Management and Sharing.” Accessed January 22. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

National Science Foundation. 2011. “Dissemination and sharing of research results.” Accessed September 3. https://www.nsf.gov/bfa/dias/policy/dmp.jsp.

O’Donnell, Angela M., and Alison King. 1999. *Cognitive Perspectives on Peer Learning*. Mahwah, N.J.: Routledge. https://search-ebscohost-com.proxy.lib.duke.edu/login.aspx?direct=true&db=nlebk&AN=24374&site=ehost-live.

Pasek, Judith E. 2017. “Historical Development and Key Issues of Data Management Plan Requirements for National Science Foundation Grants: A Review.” *Issues in Science and Technology Librarianship,* (Summer) https://doi.org/10.5062/F4QC01RP.

Piwowar, Heather A. 2011. “Who Shares? Who Doesn’t? Factors Associated with Openly Archiving Raw Research Data.” *PLoS ONE* 6 (7): e18657. https://doi.org/10.1371/journal.pone.0018657.

PLOS. 2014. “Data availability policy.” Accessed September 3. https://journals.plos.org/plosone/s/data-availability.

Poole, Alex H. 2017. “‘A Greatly Unexplored Area’: Digital Curation and Innovation in Digital Humanities.” *Journal of the Association for Information Science and Technology* 68 (7): 1772–81. https://doi.org/10.1002/asi.23743.

Rindfleish, Jennifer, Alison Sheridan, and Sue-Ellen Kjeldal. 2009. "Creating an “agora” for storytelling as a way of challenging the gendered structures of academia.” *Equal Opportunities International* 28 (6): 486-499. https://doi.org/10.1108/02610150910980783.

Schonfeld, Roger C. and Kate Wulfson. 2014. “Ithaka S+ R, Jisc, RLUK UK Survey of Academics 2012.” *Ithaka S+R*. Last Modified 14 May 2013. https://doi.org/10.18665/sr.22526.

Shaw, Patricia (1997). “Intervening in the shadow systems of organizations: consulting from a complexity perspective.” *Journal of Organizational Change Management* 10 (3): 235-50. https://doi.org/10.1108/09534819710171095.

Sigit Sayogo, Djoko, and Theresa A. Pardo. 2013. “Exploring the determinants of scientific data sharing: understanding the motivation to publish research data.” *Government Information Quarterly* 30 (Supplement 1): S19-S31. https://doi.org/10.1016/j.giq.2012.06.011.

Tenopir, Carol, Suzie Allard, Kimberly Douglass, Arsev Umur Aydinoglu, Lei Wu, Eleanor Read, Maribeth Manoff, and Mike Frame. 2011. “Data Sharing by Scientists: Practices and Perceptions.” *PLoS ONE* 6 (6): e21101. https://doi.org/10.1371/journal.pone.0021101.

Tenopir, Carol, Elizabeth E. Dalton, Suzie Allard, Mike Frame, Ivanka Pjesivac, Ben Birch, Danielle Pollock, and Kristina Dorsett. 2015. "Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide." *PLOS ONE* 10 (8): e0134826. https://doi.org/10.1371/journal.pone.0134826.

Tenopir, Carol, Natalie M. Rice, Suzie Allard, Lynn Baird, Josh Borycz, Lisa Christian, Bruce Grant, Robert Olendorf, and Robert J. Sandusky. 2015. "Data Sharing, Management, Use, and Reuse: Practices and Perceptions of Scientists Worldwide." *PLOS ONE* 15 (3): e0229003. https://doi.org/10.1371/journal.pone.0229003.

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

Weller, Travis, and Amalia Monroe-Gulick. 2014. "Understanding methodological and disciplinary differences in the data practices of academic researchers." *Library Hi Tech* 32 (3): 467–482. https://doi.org/10.1108/LHT-02-2014-0021.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. "The FAIR guiding principles for scientific data management and stewardship." *Scientific Data* 3: 160018. https://doi.org/10.1038/sdata.2016.18.

Whyte, Angus, and Graham Pryor. 2011. "Open Science in Practice: Researcher Perspectives and Participation." *International Journal of Digital Curation* 6 (1): 199–213. https://doi.org/10.2218/ijdc.v6i1.182.

Wynne, Ben, Simon Dixon, Neil Donohue, and Ian Rowlands. 2016. "Changing the Library Brand: A Case Study." *New Review of Academic Librarianship* 22 (2-3): 337-349. https://doi.org/10.1080/13614533.2016.1156000.

Vaughan, KTL, Barrie E. Hayes, Rachel C. Lerner, Karen R. McElfresh, Laura Pavlech, David Romito, Laurie H. Reeves, and Erin N. Morris. 2013. "Development of the research lifecycle model for library services." *Journal of the Medical Library Association* 101 (4): 310-314. https://doi.org/10.3163/1536-5050.101.4.013.

Zenk-Möltgen, Wolfgang, Ersa Akdeniz, Alexia Katsanidou, Verena Naßhoven, and Ebru Balaban. 2018. "Factors influencing the data sharing behavior of researchers in sociology and political science." *Journal of Documentation* 74 (5): 1053-1073. https://doi.org/10.1108/JD-09-2017-0126.