
Note: The IQ editors have granted permission for this article to be published in a special font to reinforce its presentation as a dramatic play.

An epic journey in sharing: The story of a young researcher's journey to share her data and the information professionals who tried to help

Sebastian Karcher, Sophia Lafferty-Hess¹

Abstract

Sharing data can be a journey with various characters, challenges along the way, and uncertain outcomes. These “epic journeys in sharing” teach information professionals about our patrons, our institutions, our community, and ourselves. In this paper, we tell a particularly dramatic data-sharing story, in effect a case study, in the form of a Greek Drama.² It is the quest of – a young idealistic researcher collecting fascinating sensitive data and seeking to share it, encountering an institution doing its due diligence, helpful library folks, and an expert repository. Our story has moments of joy, such as when our researcher is solely motivated to share because she wants others to be able to reuse her unique data; dramatic plot twists involving IRBs; and a poignant ending. It explores major tropes and themes about how researchers’ motivations, data types, and data sensitivity can impact sharing; the importance of having clarity concerning institutional policies and procedures; and the role of professional communities and relationships. Just like the chorus in Greek drama provides commentary on the action, a chorus of data elders in our drama points out larger lessons that the case study has for research data management and data sharing. Where actors in the Greek chorus were wearing masks, our chorus carries different items, symbolizing their message, on every entry.

Keywords

data sharing, research data management, qualitative data, collaboration

DRAMATIS PERSONAE

Jessica, A young undergraduate researcher

Sophia and Jen, Two library data folks at Duke University

Sebastian and Dessi, Two expert repository folks at QDR (Syracuse University)

Institutional Review Board staff at Duke University

The Chorus of Data Elders

PROLOGUE

Through their work, researchers produce a commodity of great value to themselves and others – data. Today, funders, journals, and research communities alike are asking researchers to share that data within a repository where it can be well cared for and made available to others (NSF, 2010; PLOS, 2014; Lupia and Elman, 2014). Sharing data in a repository has many benefits – extending the usefulness of the data past the original research question, preserving the data, enhancing transparency and reproducibility of findings, and increasing opportunities for collaboration. At Duke University, a young social science researcher has heard the call. During her undergraduate studies, she has collected interview data and after publishing on her research (Van Meir, 2017), she is now looking for help allowing others to access these unique, but sensitive, qualitative data. She has turned to the library to assist her on her journey.

ENTRY OF CHARACTERS

Duke University Libraries' Staff. Duke University Libraries provides research data management (RDM) support through their Data and Visualization Services department. RDM staff at Duke assist researchers throughout the research data lifecycle from the planning phase to sharing their data in a repository.

Qualitative Data Repository Staff. The Qualitative Data Repository (QDR) was founded at Syracuse University in 2012 to “select, ingest, curate, archive, manage, durably preserve, and provide access to digital data used in qualitative and multi-method social inquiry” (QDR, 2017). QDR is the only domain repository in the US with a sole focus on qualitative social science data. (Karcher, Kirilova, & Weber, 2016, Note 4)

Researcher. A researcher collects or generates academic data that provides the evidence for their research claims. Researchers, including students and faculty, are the primary clients of library RDM services as well as repository services.

Institutional Review Board. Institutional Review Boards (IRBs) support the ethical collection, storage, and dissemination of research involving human participants by reviewing, approving, and amending applications and protocols for research projects.

EPISODE 1

But now I will tell the lineage and the names of the heroes, and of the long sea-paths and the deeds they wrought in their wanderings; may the Muses be the inspirers of my song!

*Appolonius, The Argonautica*³

Sophia and Jen, library data folks, prepare to meet with Jessica, an undergraduate researcher. It is a beautiful day in April.

As is typical, the consult begins with Jessica discussing pertinent details on the type of data she has collected, her motivation to share, and her specific data sharing challenges. Let us examine each of these in turn. The *data* are primarily qualitative interviews with South American sex workers and NGOs. These data are not easy to come by and represent a large corpus of around 100 interviews. As far as her *motivation*, she does not have to comply with any data-sharing mandate but chooses to share. Her main

motivation is to give voice to her interviewees, whose stories are rarely listened to. She is particularly concerned with allowing others to benefit from her effort and to reuse the data in new research because as an undergraduate student she is unsure if she will be in a position to publish more on the data going forward.

However, she also faces a common *challenge* - she did not initially plan for data sharing when she wrote her IRB protocol, and therefore, did not gain consent for archiving and sharing. She has spoken to the IRB and they can help. A plan is devised where she will 1) re-consent the participants she has contact information for (around 17), 2) follow a de-identification protocol to remove all direct and indirect identifiers from the interviews, and 3) deposit the data within a protected access-restricted repository. To accomplish this she will need to file an amendment with the IRB containing details on where she will ultimately deposit her data including information on the repository security protocols, the deposit agreement, and access procedures.

Given this background, Sophia and Jen consider the situation. The plan outlined already addresses some of their common concerns about ethical data sharing. The question remains - where can Jessica deposit her data that will both comply with the IRB requirements and meet her needs? The repository must be able to provide mediated restricted access while also allowing access requests from South American researchers. Using their knowledge of domain repositories that support these types of qualitative data, they identify some options. They describe the options, promise to reach out to these repositories, and get back to the Jessica with a final recommendation for the best home for her data. Back in their offices, they reach out to contacts made through professional networks and communities (such as colleagues met at conferences), and discuss the researcher's needs. They determine a recommendation – the Qualitative Data Repository.

The CHORUS OF DATA ELDERS enters, each carrying a laptop.

INTERLUDE 1

What can information professionals learn from this episode? First, there is universal joy at the motivation to share - not based in the “stick” of mandates but the “carrot” of contributing to something larger. Understanding what affects a researcher's willingness to share provides a foundation for advocating for data sharing. While strong journal data sharing policies have been found to affect the rate of data sharing (Piwowar, 2011; Vines et al., 2013), researchers have also self-identified scholarly altruism as a driving factor amongst others (Kim and Stanton, 2015). In this story, the researcher's motivation toward reuse suggests that quantifying data reuse (Piwowar and Vision, 2013) as well as gathering a shared corpus of positive data reuse stories and exemplars could provide useful tools for advocacy and education in future. The undergraduate status of the researcher also presents opportunities to consider how age and experience impacts willingness to share. The relationship between age and data-sharing behavior is a complicated question (Tenopir et al., 2011; Tenopir et al., 2015), but understanding this relationship opens doors to target our message and services to a new generation of researchers.

Second, what of the help provided by the library data folks? Many academic institutions are increasing, or planning to increase, data management services in the face of growing need (Cox and Pinfield, 2013; Tenopir et al., 2014). Local data professionals provide a first line of support for their research community

and draw upon in-depth knowledge of data management resources and best practices. This knowledge base comes from education, hands-on experience, and active engagement with professional communities and networks. (Rice & Southall, 2016, Chap. 1) By harnessing all available resources, including resources external to one's local institution, data professionals can provide more holistic service delivery. As seen through this story, these resources also include the human relationships established through professional interactions at meetings and conferences.

EPISODE 2

Meantime from the ship the chiefs had sent Aethalides the swift herald, to whose care they entrusted their messages and the wand of Hermes, his sire, who had granted him a memory of all things, that never grew dim.

Appolonius, The Argonautica

Two weeks later, two offices, 1000 miles apart, one in the sun, the other in a snow-covered building. Library data expert Sophia reaches out to repository expert Sebastian at QDR.

After determining their recommendation, Sophia further discusses with Sebastian QDR's ability to safeguard and provide access to Jessica's data. Having ensured that QDR would be able to meet all the requirements posed by Jessica and her data, an initial meeting is organized. Some weeks later, using Skype, Jessica and Sophia meet with Sebastian and Dessi (also at QDR) to discuss the data deposit and the required next steps. Given the salience of ethical concerns in sharing qualitative data, QDR had frequently advised on IRB applications and informed consent language (see Kirilova and Karcher 2017 for some lessons) and is able to provide advice for the IRB amendment and consent language to use when re-contacting interviewees.

One open question at the outset is the right level of access controls on the data: who will be able to access the data and under what conditions? Everyone on the call shares the same two goals: to provide the greatest ease of access to the data possible while also not risking the confidentiality promised to participants. In other words, they want to make the data "as open as possible, as closed as necessary" (H2020 Programme, 2016, p. 4). Together they assess the risks from a breach in confidentiality (no risk of criminal prosecution, but potentially significant reputational harm) as well as disclosure risks. These turn out to be relatively limited, as interviews contain few indirect identifiers and the researcher is confident in her de-identification protocol. The four agree on light restrictions that require a research plan and an established academic affiliation as conditions for access. These conditions are easy to meet for interested parties and easy for the repository to investigate with no need to contact Jessica, thus ensuring access to the data in the long term.

Our four adventurers also hit on some good luck during the conversation. Qualitative data in languages other than English often poses a particular challenge because repository staff is not (or only poorly) able to check de-identification, thus removing an important safety check on the data. As luck would have it, Sebastian had conducted extensive research in some of the same areas as Jessica and would be able to check transcripts for identification risks in their original language. Moreover, QDR had recently published de-identified data from researchers working in a somewhat similar context (Dunning and Camp, 2015) who had established and published an extensive de-identification protocol, which they could point to as an example for Jessica.

The CHORUS OF DATA ELDERS enters, each wearing a headset.

INTERLUDE 2

“Sharing information,” writes Nancy Van House (2002) “requires that users and providers trust one another.” This trust is the driver of the story we are telling today. It begins with the Duke IRB and Jessica’s thesis advisor trusting an inexperienced researcher to research a sensitive topic abroad. It continues with Jessica trusting her library for advice on sharing data. It continues with Sophia and Jen trusting QDR and its staff to not just treat the data responsibly, but also to be respectful to Jessica and her expectations.

This trust goes beyond the institutional trust embodied by such notions as “trustworthy digital repository” (Beagrie et al., 2002). It relies on communication and personal interaction. It relies on a faculty advisor and an IRB taking undergraduate research seriously and being willing to take the time to guide Jessica. It takes the daily outreach of library staff to establish the library as a place that researchers – from undergraduate students to senior faculty – trust for advice on questions of data and data sharing. It takes the personal relationships among data professionals, built through meetings such as IASSIST, RDAP, and IDCC, to establish the trust between repositories and libraries that helped initiate the contact between Duke and QDR.

This trust also relies on different stakeholders playing their role throughout the data lifecycle and interacting constructively. Library data professionals are uniquely situated to serve as a point of contact for researchers, build long-term relationships, and provide in-person advice and consultations. IRBs safeguard human participants and can play an important role in helping researchers navigate the ethics of data sharing. Finally, domain repositories can provide specialized advice and data services, especially for complex data and data with complex privacy requirements. Combining data and subject-level expertise, they are also in a position to actively collaborate with researchers during appraisal (i.e., assessing whether the data are a good fit and shareable in the repository) and curation to ensure ethical sharing and to help make available high-quality data and metadata. Open lines of communication between all these stakeholders creates a “trusted network” that researchers can rely on to help them along their data journey.

EPISODE 3

Let justice and right, to which we have both agreed, stand firm.

Appolonius, The Argonautica

Back at Duke. A warm North Carolina summer.

In the weeks following the conversation, Jessica receives additional documentation from QDR with details about the proposed handling of sensitive data and suggestions for de-identification procedures as well as informed consent language. Armed with this information, she files an amendment with the Duke IRB. The IRB sends some additional inquiries about data handling and ownership, which Jessica forwards to QDR, who provides the requested information.

However, in August Jessica receives bad news from the IRB. Since she had now graduated from Duke University, and the IRB is no longer responsible for oversight of her work, so her IRB amendment will no longer be accepted. Moreover, by Duke Policy, undergraduate researchers are not able to take data containing identifiable information with them when leaving the institution (Duke Campus IRB, n.d.). Not being able to take the data with interviewees' identities means that Jessica will not be able to re-contact and re-consent her participants. The data cannot be shared.

The CHORUS OF DATA ELDERS enters, each carrying a copy of the Common Rule.

INTERLUDE 3

As we watch this epic journey unfold, let us further consider the role of the IRB in the data sharing landscape. In this story, the IRB performed its due diligence and followed the appropriate procedures to help the researcher amend the protocol for data sharing. Recognition among IRBs of the relevance of data sharing to their work, awareness of IRB's role in open science, and increasing dialogs between IRBs and data professionals will be crucial to allow for the responsible and ethical sharing of human participant data (Elman, Hoelter, Kapiszewski, and Kirilova, 2017). This growing awareness has recently been exemplified by the inclusion of explicit data sharing language in the informed consent template proposed by Cornell's IRB (Karcher, 2017). Further initiatives might also include IRBs expanding guidelines for qualitative data sharing (Jones et al., 2018) or developing joint workshops with data professionals (Duke Graduate School, 2018). These types of collaborative projects and dialogues can likewise help information professionals have clarity concerning institutional policies and procedures.

A final broad theme evoked by this story is the importance of data management and sharing education. Education about the "what" and "how" of data sharing builds awareness around good practices, resources, and tools. Different pedagogical methods have been used by the RDM community including asynchronous online trainings (EDINA, 2017), in-person workshops and trainings, data management curricula within the classroom (Whitmire, 2015), as well as data stories (DataONE, n.d.). However, conceptual education can only go so far, the actual "doing" of sharing data provides essential real world experience and lays the groundwork for future sharing. That is why these "epic journeys in sharing" are important; they are the ultimate learning exercises for both researchers and information professionals. They show us the pitfalls and possibilities for data sharing out in the wild. They build capacity and present opportunities for reflection on what worked and what went wrong.

EXODUS

Ill-starred one, why art thou so smitten with despair? We know how ye went in quest of the golden fleece; we know each toil of yours, all the mighty deeds ye wrought in your wanderings over land and sea.

Appolonius, The Argonautica

Given the eventual outcome (the inability to share the data), this story may seem tragic. Since shared data on sex workers are rare, the data would have been quite valuable, and significant investment had gone

into making it available between the researcher, Duke University Libraries, and QDR. Yet in spite of the disappointment, this did not feel like a tragedy to those of us involved. In the course of the attempted data deposit, all involved gained important knowledge – about the ethics and logistics of data sharing, about our own approaches working with researchers and IRBs, and about the particular opportunities and challenges working with undergraduate researchers.

And finally, this is *not* the end of the story. As we all know, research follows the research data *lifecycle*: the end of one research product is the beginning of the next. In early 2018, Jessica, now in graduate school, contacted Dessi for advice on consent language to support data sharing in her next research project.

THE END (for now)

ACKNOWLEDGEMENTS

The authors would like to thank their colleagues, Jennifer Darragh and Dessi Kirilova, who played a part in this sharing epic journey. Particular thanks to Jessica Van Meir, who inspired this story, gave us permission to tell it, and provided comments on a draft.

REFERENCES

- Beagrie, N., Doerr, M., Hedstrom, M., Jones, M., Kenney, A., Lupovici, C., ... Woodyard, D. (2002). Trusted digital repositories: Attributes and responsibilities (RLG-OCLC Report). Mountain View, CA: RLG. Retrieved from: <https://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>
- Cox, A., & Pinfield, S. (2014). Research data management and libraries: Current activities and future priorities. *Journal of Librarianship and Information Science* 46, 299–316. <https://doi.org/10.1177/0961000613492542>
- DataONE. (n.d.). *Data stories*. Available from: <https://www.dataone.org/data-stories>
- Dunning, T., & Camp, E. (2015). *Brokers, voters, and clientelism: The puzzle of distributive politics* [Data set]. <https://doi.org/10.5064/F6Z60KZB>
- Duke Graduate School. (2018). *RCR forum: Developing a good informed consent process*. Retrieved from: <https://gradschool.duke.edu/student-life/events/rcr-forum-developing-good-informed-consent-process>
- Duke Campus Institutional Review Board. (n.d.). *IRB Policies: Undergraduate students as researchers*. Retrieved from: <https://campusirb.duke.edu/irb-policies/undergraduate-students-researchers>
- EDINA. (2017). MANTRA: Research data management training. University of Edinburgh. Retrieved from: <https://mantra.edina.ac.uk/>
- Elman, C., Hoelter, L., Kapiszewski, D., & Kirilova, D. (2017, November). *IRB Guidelines and Data Sharing in the Social Science: Tensions and Strategies to Address Them*. Presented at the PRIM&R Social, Behavioral, and Educational Research Conference, San Antonio, TX. <https://doi.org/10.6084/m9.figshare.5969104.v1>
- 7/9 Karcher, Sebastian and Sophia Lafferty-Hess (2019) An epic journey in sharing: The story of a young researcher's journey to share her data and the information professionals who tried to help, IASSIST Quarterly 43 (1), pp. 1-9. DOI: <https://doi.org/10.29173/iq942>

- H2020 Programme. (2016). *Guidelines on FAIR Data Management in Horizon 2020*. Brussels: European Commission Directorate for General for Research & Innovation. Retrieved from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- Jones, K., Alexander, S.M., Bennett, N., Bishop, L., Budden, A., Cox, M., ... Winslow, D. (2018). Qualitative data sharing and re-use for socio-environmental systems research: A synthesis of opportunities, challenges, resources and approaches. Digital Repository at the University of Maryland. <https://doi.org/10.13016/M2WH2DG59>
- Karcher, S. (2017, February 9). Participant Protection, Informed Consent, and Data Sharing. Retrieved March 21, 2018, from <https://qdr.syr.edu/qdr-blog/participant-protection-informed-consent-and-data-sharing>
- Karcher, S., Kirilova, D., & Weber, N. (2016). Beyond the matrix: Repository services for qualitative data. *IFLA Journal*, 42(4), 292–302. <https://doi.org/10.1177/0340035216672870>
- Kirilova, D., & Karcher, S. (2017). Rethinking data sharing and human participant protection in social science research: Applications from the qualitative realm. *Data Science Journal*, 16. <https://doi.org/10.5334/dsj-2017-043>
- Lupia, A., & Elman, C. (2014). Openness in political science: Data access and research transparency. *PS: Political Science & Politics*, 47(1), 19–42. <https://doi.org/10.1017/S1049096513001716>
- National Science Foundation (NSF). (2010). *Dissemination and sharing of research results*. Arlington, VA: National Science Foundation. Retrieved from: <https://www.nsf.gov/bfa/dias/policy/dmp.jsp>
- PLOS. (2014). *Data availability policy*. Retrieved from: <http://journals.plos.org/plosone/s/data-availability>
- Piwowar, H.A. (2011). Who shares? Who doesn't? Factors associated with openly archiving raw research data. *PLoS ONE* 6, e18657. <https://doi.org/10.1371/journal.pone.0018657>
- Piwowar, H.A., & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ* 1:e175 <https://doi.org/10.7717/peerj.175>
- Qualitative Data Repository (QDR). (2017). *Our mission*. Retrieved from: <https://qdr.syr.edu/>
- Rice, R., & Southall, J. (2016). *The data librarian's handbook*. London: Facet Publishing. Abgerufen von <http://www.facetpublishing.co.uk/title.php?id=300471>
- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A.U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLoS ONE* 6, e21101. <https://doi.org/10.1371/journal.pone.0021101>
- Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLOS ONE* 10, e0134826. <https://doi.org/10.1371/journal.pone.0134826>

- Tenopir, C., Sandusky, R.J., Allard, S., & Birch, B. (2014). Research data management services in academic research libraries and perceptions of librarians. *Library & Information Science Research* 36, 84–90. <https://doi.org/10.1016/j.lisr.2013.11.003>
- Vines, T.H., Andrew, R.L., Bock, D.G., Franklin, M.T., Gilbert, K.J., Kane, N.C., Moore, J.-S., Moyers, B.T., Renaut, S., Rennison, D.J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal* 27, 1304–1308. <https://doi.org/10.1096/fj.12-218164>
- Van Meir, Jessica. (2017). Sex work and the politics of space: Case studies of sex workers in Argentina and Ecuador. *Social Sciences* 6 (2): 42. <https://doi.org/10.3390/socsci6020042>.
- Whitmire, A.L. (2015). Implementing a graduate-level research data management course: Approach, outcomes, and lessons learned. *Journal of Librarianship and Scholarly Communication* 3(2), peP1246. <https://doi.org/10.7710/2162-3309.1246>

¹ Both authors have contributed equally to this paper and are listed alphabetically. Sebastian Karcher, Associate Director, Qualitative Data Repository, skarcher@syr.edu. Sophia Lafferty-Hess, Research Data Management Consultant, Duke University Libraries, sophia.lafferty.hess@duke.edu.

² The narrative structure of this paper was inspired by the IASSIST 2018 conference theme of “Once Upon a Data Point: Sustaining Our Data Storytellers.”

³ All quotes from *The Argonautica*, Appolonius’ ancient account of a band of heroes on an improbable quest, are from the English translation by R. C. Seaton available on the *Internet Classics Archive* at <http://classics.mit.edu/Apollonius/argon.html>