

Perceptions and Experiences of Qualitative Open Data (PEQOD): Exploratory Pilot

Michelle M. Kazmer¹

¹ Florida State University, Tallahassee, FL 32306, USA
mkazmer@fsu.edu

Abstract. This study examines open qualitative human subjects research data and the experiences with open data policies of researchers who use this type of data. The goal of the overall project will be to generate empirical findings and use them to develop a conceptual and practical framework to help researchers and institutions frame their open data practices for qualitative human subjects data within existing technical systems and in accordance with legal, institutional, funding agency, and publisher requirements. This poster frames the problem of qualitative open data in the context of existing literature and policies and reports the findings of an exploratory pilot study comprising semi-structured interviews undertaken with six qualitative researchers from six disciplines. As the participants considered their own qualitative human subjects data in the context of open data policies, the concepts that emerged include: the situated, co-constructed, non-neutral nature of qualitative human subjects data; ethical obligations and logistical arguments for and against re-use of these data; the “stakes” or implications of the content of the data and its confidentiality; and metadata to support ethical and effective data re-use.

Keywords: Qualitative Research, Open Data, Research Data Management

1 Introduction

1.1 Qualitative Open Data

An emerging area within the creation/curation of research data is *open qualitative human subjects research data* and the experiences with *open data policies* of researchers who collect and use this data. *Qualitative human subjects data* are diverse and vary in many ways including content, collection type, and storage format. *Open data policies* are requirements or affirmations by funding agencies, publishers, academic associations, and research institutions for research data to be made openly available, often through deposit in an online data repository [1].

Because of the lack of information available in the existing literature it is not clear how much *researchers who collect and use qualitative data* know about open data policies; how they perceive those policies as applying to their data; what experience they have in making their data available openly; and what factors may facilitate or hinder their participation in open data initiatives [2, 3, 4].

1.2 Open Data Background

The need for research data to be made publicly available has been stated explicitly by funding agencies, research institutions, and publication venues. The U. S. National Science Foundation has a suite of documents affirming the need for data to be available, accessible, and usable [5], and the U. S. National Institutes of Health have a similar position [6]. In 2013 the U.S. Office of Science and Technology Policy (OSTP) issued a policy memorandum on access to research findings [7], and in 2017 the OSTP signed the new Executive branch's 2019 priorities which include a focus on storage and re-use of research findings [8]. The Association of American Universities affirms the importance of making the results of research open to the public [9], and our own institution (blinded) has like many others developed a public access policy for research data. Open-access research journals such as those published by PLOS require researchers to make their data available openly; highly-ranked journals such as *Nature* and *Science* have similar requirements.

Most data shared under such requirements are quantitative data, collected not from human subjects or, if collected from human subjects, fairly easily anonymized without compromising the utility of the data [10, 11, 12]. Even when quantitative data are collected from human participants and subject to protections through regulatory instruments (in the U.S.) covering health data and research participant confidentiality, there are an increasing number of institutional guidelines and tools for making the data available while maintaining regulatory compliance [13, 14, 15]. There are fewer direct discussions of, or tools to support, qualitative researchers and the specific characteristics and needs of their data [16], although several recent studies have used qualitative approaches to study open data practices with respect to quantitative data in varied contexts [17, 18, 19].

2 Research Questions and Methods

This research used an emergent thematic approach to data collection and analysis and is designed to answer the following exploratory research questions: **RQ1:** How do the qualitative researchers in this study describe their understanding and awareness of open data policies? **RQ2:** What experiences do the qualitative researchers in this study have with open data policies and practices? **RQ3:** How do the qualitative researchers in this study think open data policies do/do not affect their research data? **RQ4:** In what ways do the qualitative researchers in this study say they believe that open data as a concept is or is not relevant to qualitative research data in general?

2.1 Recruiting and participants

During Spring and Summer 2017, the author designed and tested a data collection instrument and ran a pilot research project to test the methods of recruiting participants and hone the data collection instrument by interviewing six qualitative researchers in six disciplines. Purposive sampling among faculty members from the author's home academic institution yielded 6 usable interviews with qualitative researchers from 6 disciplines: communication; criminology; education; family & child sciences; information science; and sport management. The participant solicitation protocol and data collection instruments were approved by the author's home university's Institutional Review Board.

2.2 Data collection instrument

The poster will include the full data collection instrument, which was omitted here to allow more details about the findings. The semi-structured interview schedule includes 19 questions (two of which have yes/no probe trees for researchers who do and do not have experience with open data). Interview questions mapped to the 4 RQs about open data: 5 questions focus on RQ1 (understanding/awareness); 5 questions focus on RQ2 (experience); 5 questions focus on RQ3 (personal relevance); and 4 questions focus on RQ4 (general relevance).

3 Analysis and Preliminary Findings

Analysis of the data collected during this exploratory study began immediately upon collection and has continued throughout and beyond the completion of the 6 interview, with the goal of building initial concepts to inform subsequent data collection and to help the author quickly understand key concerns and opportunities in the area of qualitative open data practices. The initial findings are under review by the research participants for validation and editing and the reviewed version will be presented in the final poster if accepted. As the researchers who were interviewed considered their own qualitative human subjects data in the context of open data policies, four initial concepts emerged:

1. The situated, co-constructed, non-neutral nature of qualitative human subjects data. The 6 researchers all focused on the context in which their data were collected and created, stressing that the meaning of the data was tied to the time, place, and circumstances in which observations and interviews occurred. Several researchers focused on the co-construction of data, stating that the data existed only through the interactions between them and their research participants. All 6 researchers noted the non-neutrality of qualitative data (some acknowledging that quantitative data are also not neutral) and agreed that, while it holds value, it cannot be treated as a fungible commodity.

2. Ethical obligations and logistical arguments for and against re-use of these data. Ethical obligations associated with informed consent and institutional review were a key concern. Most of the researchers went further, indicating that their ethical concerns about confidentiality, downstream identifiability by data aggregators, and mis-use of the data by others went well beyond what was required by institutional review and informed consent procedures.
3. The “stakes” or implications of the content of the data and its confidentiality. Related to, but distinct from, the ethical obligations of identifiability and confidentiality, were concerns about the difference in “stakes”, implications, repercussions or unintended consequences of making qualitative human subjects data openly available in a repository from which it could be retrieved by “anyone” and used either thoughtlessly, without consideration to the original participants, or maliciously, by people intending to do harm to the original participants, to the institution, or to the research arena at large. The researchers acknowledged the importance of research replicability and evaluation, but most were extremely skeptical of the possibility that someone else could pick up their dataset and re-analyze it meaningfully.
4. Metadata to support ethical and effective re-use. Only one participant in this pilot was from information science and the other five did not use the term metadata. All six did, though, wrestle with the idea of how to have some extra information travel with the data to help explain to different audiences what it was, how it was collected, the theories, models, and/or research questions that informed its collection and initial analysis, and how it should be thought about in a potential re-use scenario.

4 Conclusion and Further Directions

It is difficult to draw firm conclusions from an exploratory pilot, but it can tentatively be asserted, subject to additional data collection, that while qualitative researchers are convinced of the ethical imperative to share their data, they are also wary of how and by whom the data may be re-used, concerned about their obligations to the participants in their research, and in many cases unconvinced that the data can be understood out of the context in which it was originally collected and analyzed.

The near-term future directions of this research are two-fold: (a) to use the now-tested interview instrument to interview 25-30 additional qualitative researchers from multiple disciplines and settings (including non-academic) outside the author's home institution; and (b) to conduct a systematic policy review of existing open data policies with a focus on the ways they are relevant, and potentially relevant, to qualitative human subjects data and with the goal of developing policy statements or modules that can be implemented within existing institutional and technical infrastructures.

References

1. Borglund, E., Engvall. T. Open data? *Records Management Journal*, 24(2), 163-180 (2014).

2. Cummings, J. A., Zagrodny, J. M., Day, T. E. Impact of open data policies on consent to participate in human subjects research: Discrepancies between participant action and reported concerns. *PLOS ONE* 10(5):e0125208 (2015).
3. Fielding, N. G. Qualitative research and our digital futures. *Qualitative Inquiry*, 20(9), 1064-1073 (2014).
4. Hossain, M. A., Dwivedi, Y. K., Rana, N. P. State-of-the-art in open data research: Insights from existing literature and a research agenda. *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 14-40 (2016).
5. National Science Foundation. Open data at NSF. <https://nsf.gov/data/>
6. National Institutes of Health. *NIH sharing policies and related guidance on NIH-funded research resources*. <https://grants.nih.gov/policy/sharing.htm>
7. U.S. Office of Science and Technology Policy 2013 memorandum <https://obamawhitehouse.archives.gov/>
8. U.S. Office of Science and Technology Policy 2017 FY 2019 priorities <https://www.whitehouse.gov/>
9. Association of American Universities <https://www.aau.edu/issues/publicopen-access>
10. Childs, S., McLeod, J., Lomas, E., Cook, G. Opening research data: Issues and opportunities. *Records Management Journal*, 24(2), 142-162 (2014).
11. DeLyser, D., Sui, D. Crossing the qualitative-quantitative chasm III: Enduring methods, open geography, participatory research, and the fourth paradigm. *Progress in Human Geography*, 38(2), 294-307 (2014)
12. Faniel, I. M., Kriesberg, A., Yakel, E. Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*, 67(6), 1404-1416 (2016).
13. Georgia Institute of Technology, <http://d7.library.gatech.edu/research-data/home>.
14. University of Michigan <https://www.icpsr.umich.edu/icpsrweb/content/datamanagement/ostp.html>
15. University of Wisconsin-Madison, <http://researchdata.wisc.edu/category/sharing-data/>
16. Corti, L., Fielding, N. Opportunities from the digital revolution: Implications for researching, publishing, and consuming qualitative research. *SAGE Open*, 6(4), 2158244016678912 (2016).
17. Bezuidenhout, L. M., Leonelli, S., Kelly, A. H., Rappert, B. Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy*, 44(4), 464-475 (2017).
18. D'Agostino, M., Mejía, F., Marti, M., Novillo, D., de Cosio, F. G., Farach, N. Social dialogue and scientific production on big and open data in health: From facilitating surveillance and preventing disease to fostering behavioral changes. *International Journal of Public Health Research*, 4(3), 14 (2016).
19. Ingrams, A. The legal-normative conditions of police transparency: A configurational approach to open data adoption using qualitative comparative analysis. *Public Administration*, 95(2), 527-545 (2017).