

Untangling Data Sharing and Reuse in Social Sciences

Renata Curty

Universidade Estadual de
Londrina
renatacurty@uel.br

Ayoung Yoon

Indiana University Purdue
University Indianapolis
ayyoon@iupui.edu

Wei Jeng

University of
Pittsburgh
wej9@pitt.edu

Jian Qin

Syracuse
University
jqin@syr.edu

ABSTRACT

The discipline of social science has unique research norms and cultures regarding data sharing and reuse that can be affected by complex factors related to context, time and dependence on human subjects. Compared with STEM disciplines, social sciences emphasize the protection of study participants and observees. Extra effort is required from reusers to preserve data interconnectedness in order to guarantee the data's understandability and informative value. In this panel, the panelists will present their research findings and provide perspective on social science data sharing and reuse, including factors that may influence data reuse behavior, researchers' trust judgment in data for data reuse, and infrastructural barriers and incentives for data sharing among social scientists. This panel aims to provide an overview of the current state of social science data reuse and sharing, and, in collaboration with panel participants, elicit topics for future research. It also proposes a practical agenda to develop alternative incentives for individual researchers, and potential ways in which data sharing and reuse can be improved, coordinated, and encouraged among social scientists.

Keywords

Data sharing, data reuse, social science, disciplinary data, data infrastructure.

INTRODUCTION

In the past decade, open science initiatives have paved the way for increased transparency and reproducibility in research. In the quest to expand the availability of research data and to comply with new governmental directives, a number of funding agencies, journals, and research organizations have started implementing data sharing mandates. Meanwhile, the scientific community has witnessed a recent propagation of research data repositories (Marcial & Hemminger, 2010). This trend enables research data to be shared and more accessible purposes which are not always anticipated by original collectors.

Despite these advances, previous studies have revealed that there are still barriers for sharing and reusing data. Such barriers include insufficient time, amount of involved effort, perceived risk of sharing or reusing others' data, data quality concerns, and lack of incentives (Tenopir et al., 2011; Kim & Zhang, 2015). These barriers negatively affect

both STEM and social science researchers alike. However, there is a lack of research on data sharing and reuse practices outside the STEM domain.

Data sharing and reuse issues in the social sciences are different from many other disciplines because of the complex contextual nature of data and its discipline-specific research practices. Social science data are fundamentally different from other scientific data that are field- or lab-based, due to human involvement and its time- and context-dependent nature. Because the majority of social science data involve direct or indirect interactions with human subjects during the data collection process, there are many proposed ethical concerns about sharing and reusing this data, particularly regarding qualitative data (e.g., Bishop, 2009; Carusi & Jirotko, 2009). As human participants are involved in social science research, ensuring confidentiality and anonymity (i.e., protecting the identity of the study participants) is critical when archiving and sharing data. Concerns about data quality can also hinder researchers from using others' data. While protecting data quality is important across disciplines, previous research suggests that data quality might carry specific weight in the social sciences (Yoon, 2016). Due to the unique characteristics of social science data, it is necessary to exclusively examine data practices in the social science setting.

While recent research has investigated social science data sharing and reuse (e.g., Faniel, Kriesberg, & Yakel, 2015; Yoon, 2014b; Kim & Adler, 2015), a number of methodological and theoretical issues could be expanded upon for a more comprehensive understanding. The first issue is related to the comparative approach to understand data sharing and reuse depending on the different types of social science data (e.g., quantitative and qualitative). For instance, qualitative data is considered to be more complex and difficult to share than quantitative data (Bishop, 2007, 2009; Coltart, Henwood & Shirani, 2013; Corti, 2007; Grinyer, 2009; Hammersley, 2010; Heaton, 2004, 2008). On one hand, from the data sharers' perspective, qualitative data requires scrupulous handling with regard to the preparation, licensing, consent, and access rights during research before it is included in a data repository. Examples of this include the anonymization of personal details and ensuring consent for sharing and potential reuse (Schäfer et al., 2011). On the other hand, from the data reusers'

perspective, qualitative data are expected to require more effort in interpreting, re-assessing, and reusing and, therefore, are less likely to be reused (Niu, 2009). Along these lines, considering the specifics of each type of data is necessary in order to promote action and provide guidance that will compel social scientists' data sharing and reusing practices.

Another issue is the holistic and multidimensional approach to study data sharing and reuse in regarding how individual, institutional, disciplinary (community), and infrastructure factors come into play in social science data sharing and reuse, which accounts for a range of significant factors. Past studies reveal the complex nature of data practices, which are affected by multiple individual (i.e., social scientists who are handling data), institutional (i.e., organizations that support or influence individuals), disciplinary, and infrastructure factors (Fitzgerald et al., 2007; Kowalczyk & Shankar, 2011; Wallis, Rolando, & Borgman, 2013). While data sharing and reuse can be a researcher's choice depending on individual need and preference (Niu, 2009; Pienta, Alter, & Lyle, 2010) and data exchange occurs person-to-person as an informal practice (Sands, Borgman, Wynholds, & Traweek, 2012), researchers' data sharing and reusing are influenced by disciplinary norms and institutional practices, similar to their day-to-day research activities (Carlson & Bracke, 2013; Elman, Kapiszewski, & Vinuela, 2010; Kim, 2007; Kim & Zhang, 2015). Previous studies also confirm that data practices and behaviors may vary due to the disciplinary or scholarly communities' influences (e.g., Birnholtz & Bietz, 2003; Carlson & Anderson, 2007; Faniel, Barrera-Gomez, Kriesberg, & Yakel, 2013; Rolland & Lee, 2013) because data sharing and reuse are an ingrained practice in some disciplinary traditions (Borgman et al., 2012; Faniel & Jacobsen, 2010). Infrastructure for data sharing and reuse also influences the researchers' data practices, since social science has a long history of supporting data sharing and reuse through data repositories and relevant policies (Cragin, Palmer, Carlson, & Witt, 2010; Daniels, Faniel, Fear, & Yakel, 2012; Yakel, Faniel, Kriesberg, & Yoon, 2013; Yoon, 2014a). Thus, employing a holistic and multidimensional approach is critical for understanding the complexity of data practices and the factors that drive or hinder data sharing and reuse.

In summary, employing a holistic and multidimensional approach is critical for understanding the complexity of data practices and factors that facilitate or hinder data sharing and reuse. The goal of this panel is to discuss the findings of research and issues and to promote social science data sharing and reuse, the detailed objectives and structures of which are described below.

OBJECTIVES OF THE PANEL

The objectives of this panel include:

- To examine the nature of data sharing and reuse specific to the social-science context;

- To discuss the fundamental characteristics of different types of social-science data (e.g., quantitative and qualitative) and their impact on data sharing and reuse;
- To explore how individual, institutional, disciplinary (or community), and infrastructure factors influence data sharing and reuse in social sciences;
- To collaboratively identify opportunities for further research.

STRUCTURE OF THE PANEL

The structure of the proposed panel is as follows:

- The moderator (Qin) will introduce the panel theme, core definitions, and elaborate the purposes and objectives of the panel in the first ten minutes.
- Each one of the panelists (Curty, Yoon, and Jeng) will present (for 10-15 minutes) their empirical research findings about data reuse and data sharing in the social sciences.
- The moderator will facilitate a group session after the panelists' presentations. The discussion will be based on a set of pre-developed questions to accomplish the objectives described above. The outcome of the discussion will assist the development of a collaborative agenda for future research and practical initiatives.

PANELISTS AND TOPICS

Each panelist brings their unique expertise to the topic of data sharing and reuse in the social sciences, as follows.

Data Reuse in Social Science

Renata Curty is an assistant professor in the Information Science Department at the State University of Londrina (Universidade Estadual de Londrina), Brazil. Her research relates to scholarly communication, data curation and research data reuse. Her current research project investigates enhanced publications and their impact on actual data reuse.

In this panel, Curty will present the findings of a mixed-method study which identify a collection of factors as influential on data reuse behaviors among social scientists. She will start by introducing the research model comprised of 25 factors classified into six main categories, which was developed based on the triangulation of her qualitative study data analysis with the literature. Then, her presentation will focus on the results of a large-scale study that, not only validated the research model, but also measured data reuse intention and actual behavior among U.S. social scientists randomly sampled from the Community of Science (CoS) database. Curty will demonstrate the factors responsible for hindering or driving data reuse intention and actual behavior among members of the surveyed population. She will also address some

behavioral differences depending on the type of data intended to be reused (i.e., quantitative, qualitative or mixed), as well as sub-disciplines.

Trust Issues in Data Reuse

Ayoung Yoon is an assistant professor at the School of Informatics and Computing at Indiana University Purdue University Indianapolis. She has completed research on data reuse, data curation, and data repositories.

Her presentation in this panel will focus on the issue of trust in data reuse among quantitative social scientists. Reusers' trust is fundamental for data to be reused, and trust is a useful theoretical concept to apply in order to understand data reusers' thoughts, experiences, and needs, as trust is woven into the life cycle of data. In this panel, she will discuss:

- How data reusers identify the trust attributes of data from their reuse experiences;
- How data reusers' trust attributes are associated with multidimensional layers (individual, institutional, disciplinary/community, and infrastructure).

The discussion will demonstrate the dynamic and social nature of data and data reuse in social science, as data reusers' trust is associated with multiple attributes rather than one, and their trust judgment occurs at various levels, including that of the object (data), individuals, institutions, communities, and society.

Data Sharing in Social Science

Wei Jeng is a Ph.D. candidate at the School of Information Sciences (iSchool) at the University of Pittsburgh. Her research explores how academics share information, data, and resources in the digital age. Given the increasing need for academic communities to manage an enormous amount of data, her long-term research goal is to provide insights for improving research infrastructure for scholars in all disciplines, particularly the social sciences, humanities, and related scholarly communities. Her working dissertation project investigates the determining factors and motivations of social scientists' data-sharing practices, especially those dealing with qualitative data in social sciences.

In previous studies Jeng recognizes four key aspects in social science data sharing: individuals, organization, discipline community, and infrastructure (Jeng, He, & Oh, 2016); being the last key dimension, the focus of her panel presentation will engage in a discussion about the *infrastructure* aspect. Using Interuniversity Consortium for Political and Social Research (ICPSR)---a primary data repository for social sciences---as a case study, Jeng will share her findings regarding the sharing and curation process in a data repository based on the surveys she conducts with the ICPSR data depositors and curators.

- For data depositors, Jeng will highlight the determinants that influence individual social scientists' behaviors when sharing their qualitative and mixed method data. This is unveiled by surveying data depositors who have shared mixed method data or qualitative data at ICPSR.
- For data curators, Jeng will discuss the findings based on a focus groups that she conducted with ICPSR directors and staff. She is particularly interested in data curation professionals' view on the challenges of curating qualitative and quantitative social science data.

In summary, Jeng aims to present her viewpoints regarding qualitative data depositors and curators, demonstrating the tension and synergy between individuals, context (institution and discipline community) and infrastructure (data repository).

The moderator

Jian Qin is a professor at the School of Information Studies, Syracuse University. She specializes in metadata, knowledge modeling and organization, research data management, and scientific communication, and has been widely published in library and information science journals. Her research has been funded by the National Science Foundation (NSF), the Institute for Museum and Library Services (IMLS), and the Interuniversity Consortium for Political and Social Research (ICPSR) and Alfred P. Sloan Foundation. Her current projects include a large scale of data mining in the GenBank data repository and creating a metadata model for gravitational wave research data management, both funded by NSF.

EXPECTED CONTRIBUTION

This panel aims to contribute to the discussion on social science data sharing and reuse in the following ways. From a research standpoint, the panelists anticipate that their empirical findings will provide participants with a more comprehensive and updated conceptual framework for understanding social scientists' perceptions and behaviors towards openly sharing their data, as well as reusing others' data. From a practical angle, this panel is expected to be valuable to librarians, policymakers, open data advocates, and data repository stakeholders to reflect on data sharing and reusing practices. This not only serves as a foundation to build more sustainable disciplinary data infrastructures, but can also facilitate data openness and collaboration in the social sciences.

REFERENCES

- Birnholtz, J. P., & Bietz, M. (2003). Data at Work: Supporting Sharing in Science and Engineering. *Proceedings of ACM Conference on Supporting Group Work*, 339-348.

- Bishop, L. (2007). A reflexive account of reusing qualitative data: Beyond primary/secondary dualism. *Sociological Research Online*, 12(3), 2.
- Bishop, L. (2009). Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues*, 44(3), 255-272.
- Borgman, C. L. (2009). The digital future is now: A call to action for the humanities. *Digital Humanities Quarterly* (3)4. Retrieve from digitalhumanities.org/dhq/vol/3/4/000077/000077.html
- Carusi A. & Jirotko M. (2009). From data archive to ethical labyrinth. *Qualitative Research*, 9(3), 285-298.
- Carlson, J., & Bracke, M. (2013). Data management and sharing from the perspective of graduate students: An examination of culture and practice at the water quality field station. *Portal: Libraries and the Academy*, 13(4), 343-361.
- Carlson, S., & Anderson, B. (2007). What Are Data? The Many Kinds of Data and Their Implications for Data Re-Use. *Journal of Computer-Mediated Communication*, 12(2), 635-651. <http://doi.org/10.1111/j.1083-6101.2007.00342.x>
- Coltart, C., Henwood, K. & Shirani, F. (2013). Qualitative secondary analysis in austere times: Ethical, professional and methodological considerations. *Historical Social Research*, 38(4), 271-292.
- Corti, L. (2007). Re-using archived qualitative data—where, how, why? *Archival Science*, 7(1), 37-54.
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038.
- Daniels, M., Faniel, I., Fear, K., & Yakel, E. (2012). Managing fixity and fluidity in data repositories. *Proceedings of the 2012 iConference* (pp. 279-286).
- Elman, C., Kapiszewski, D., & Vinuela, L. (2010). Qualitative data archiving: Rewards and challenges. *Political Science & Politics*, 43(1), 23-27.
- Faniel, F., Barrera-Gomez, J., Kriesberg, A., & Yakel, E. (2013). A Comparative study of data reuse among quantitative social scientists and archaeologists. *Proceedings of iConference 2013*, 797-800.
- Faniel, I. M., Kriesberg, A., & Yakel, E. (2015). Social scientists' satisfaction with data reuse. *Journal of the Association for Information Science and Technology*.
- Faniel, I. M., & Jacobsen, T. E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work*, 19(3-4), 355-375.
- Fitzgerald, A. M., Pappalardo, K. M., Fitzgerald, B. F., Austin, A. C. ..., Singleton, B. (2007). Building the infrastructure for data access and reuse in collaborative research: An analysis of the legal context. *The OAK Law Project*, Canberra, Australia.
- Grinyer, A. (2009). The ethics of the secondary analysis and further use of qualitative data. *Social Research Update*, 56(4), 1-4.
- Hammersley, M. (2010). Can we re-use qualitative data via secondary analysis? Notes on some terminological and substantive issues. *Sociological Research Online*, 15(1), 5
- Heaton, J. (2004). *Reworking Qualitative Data*. Thousand Oaks, CA: SAGE Publications.
- Heaton, J. (2008). Secondary analysis of qualitative data: An overview. *Historical Social Research/Historische Sozialforschung*, 33-45.
- Jeng, W., He, D., & Oh, J. (2016). Toward a conceptual framework for data sharing practices in social sciences: A profile approach. *Proceedings of the ASIS&T 2016 Annual Meeting*.
- Kim, J. (2007). Motivating and impeding factors affecting faculty contribution to institutional repositories. *Journal of Digital Information*, Vol 8. Retrieved from <http://journals.tdl.org/jodi/article/view/193/177>
- Kim, Y., & Adler, M. (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.
- Kim, Y., & Zhang, P. (2015). Understanding data sharing behaviors of STEM researchers: The roles of attitudes, norms, and data repositories. *Library & Information Science Research*, 37(3), 189-200.
- Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual Review of Information Science and Technology*, 45(1), 247-294.
- Marcial, L. H. & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10), 2029-2048.
- Niu, J. (2009). Overcoming Inadequate Documentation. *Proceedings of the Annual Meeting of the American Society for Information Science & Technology*.
- Pienta, A. M., Alter, G. C., & Lyle, J. A. (2010). The enduring value of social science research: The use and reuse of primary research data. *The Organisation, Economics and Policy of Scientific Research Workshop*. Retrieved from <http://deepblue.lib.umich.edu/handle/2027.42/78307>
- Rolland, B., & Lee, C. P. (2013). Beyond trust and reliability: reusing data in collaborative cancer epidemiology research. *Proceedings of the 2013 Conference on Computer supported cooperative work*, 435-444.
- Sands, A., Borgman, C. L., Wynholds, L., & Traweek, S. (2012). Follow the data: How astronomers use and reuse data. *Proceedings of the American Society for Information Science and Technology*, 49(1), 1-3.

- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., ... Frame, M. (2011). Data sharing by scientists: Practices and perceptions. *PLOS ONE*, 6(6), e21101.
- Yakel, E., Faniel, I., Kriesberg, A., & Yoon, A. (2013). Trust in Digital Repositories. *International Journal of Digital Curation*, 8(1), 143–156.
- Yoon, A. (2014a). End users' trust in data repositories: definition and influences on trust development. *Archival Science*, 14(1), 17–34.
- Yoon, A. (2014b). "Making a square fit into a circle": Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4.
- Yoon, A (2016). Visible evidence of invisible quality dimensions and the role of data management. *Proceedings of iConference 2016*.
- Yoon, A. (in press). Data reusers' trust development. *Journal of the Association for Information Science and Technology*.