
The Ethos and Pragmatics of Data Sharing

Ingrid Erickson

School of Communication and Information
Rutgers University
ingrid.erickson@rutgers.edu

Kristin Eschenfelder

School of Library & Information Studies
University of Wisconsin, Madison
eschenfelder@wisc.edu

Sean Goggins

School of Information Science and Learning Technologies
University of Missouri
gogginsp@missouri.edu

Libby Hemphill

Communication and Info Studies
Illinois Institute of Technology
libby.hemphill@iit.edu

Steve Sawyer

School of Information Studies
Syracuse University
ssawyer@syr.edu

Kalpana Shankar

School of Info & Library Studies
University College Dublin
kalpana.shankar@ucd.ie

Katie Shilton

College of Information Studies
University of Maryland
kshilton@umd.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).
CSCW'14 Companion, February 15–19, 2014, Baltimore, MD, USA.
ACM 978-1-4503-2541-7/14/02.
<http://dx.doi.org/10.1145/2556420.2556852>

Abstract

The focus of this panel is the pragmatics of data sharing as framed by the needs and pressures of scholarly work. Panelists represent a lively blend of quantitative, qualitative and mixed methods researchers with recent experiences in developing and sharing data. Panelists will present research and address questions related to data collection and management, human subjects protocols, data archival and data repositories and other emergent issues.

Author Keywords

Data sharing, data set, collaboration, infrastructure, archive, data management

ACM Classification Keywords

K.4.0. Computers in Society: general.

Introduction

It has become increasingly clear that data—primarily, but not exclusively, digital data—have taken on a new value both in commerce and research in recent years. Long-prized in the past as a source for gleaning insight on discrete projects within delimited domains, data are now being seen with equal esteem as potential inputs to larger, systems-level analyses made possible by increased computing power and interlinked infrastructures.

In line with this shift, we have seen the development of a new ethos of openness and collaboration that reflects the dual logics of increased power via aggregation and federation, and increased virtue via transparency and accessibility. This ethos has facilitated the development of new tools, strategies and mandates to extract, archive and share data that push against established protocols and conventions regarding anonymity, privacy, confidentiality, security and other values in both qualitative and quantitative research.

Moreover, whether welcomed or resisted, data sharing requires that researchers develop new ways of conducting their research. Questions such as the following demand engagement:

- How should data be collected, named, and archived by researchers who aim to share data sets with others? How might data provenance be articulated?
- How should subjects be informed that their data might be available to others? How might subjects' rights be upheld in this era of openness and accessibility? How might IRB protocols need to adapt?
- When should data be shared? And for how long into the future?
- How should data sets be made accessible to others and in what formats? What types of repositories or other exchange mechanisms should be developed to accommodate this access? How should these mechanisms be managed?

- How should we consider data sharing with regard to work/labor issues? Whose job is this? Are data management and data sharing practices sustainable?
- How do institutionalized computer-supported data sharing practices compare to ad hoc/peer-to-peer negotiated practices?

Pragmatic solutions to these and other salient questions are currently being worked out by researchers, but as of yet no particular methods or processes have become institutionalized. This lack of a coherent, multi-disciplinary strategy for data sharing is itself another interesting facet of the larger conversation.

Our panel brings together a set of scholars with interests and expertise in data practices [1][2][3][4][5][6][7] to discuss several of the current opportunities and complexities involved with data sharing and management, both from the perspective of applied practitioners as well as ethical researchers. Our goal with this panel is to address all of the questions above and to elicit conversation from attendees that begins to shape an agenda for a future publication or a future panel on a related theme(s).

We envision the following panel format can facilitate both knowledge exchange and discussion as we intend:

- Short introductory talks from each panelist that identify the context for the discussion to follow. We will ask panelists to speak to a specific set of prompts, including a description of data type(s), data management strategies, and insights gained from past data sharing

experiences. We expect these talks to be quick and informational, not longer than 5-7 minutes ideally.

- During the panelist presentations, we will distribute cards among attendees so that they have the opportunity to ask specific questions of panelists or make comments as desired.
- The session organizer will collect and organize comments and questions so that panelists can address them thematically, as a group, for the remainder of the session. This process is meant to support public discussion among panelists as well as providing questioners and commentators with direct feedback.

Panelists

Kristin Eschenfelder

Eschenfelder is a Professor and Director at the School of Library and Information Studies at the University of Wisconsin-Madison. Her research interests focus on access and use regimes – or the complex, multi-level networks of laws, customs, technologies and expectations that shape what information we can access in our daily lives and how we can make of it. Her recent work has compared data access and use rules in data repositories across the social sciences, humanities, biological sciences and physical sciences. With Shankar, she is currently conducting a case study of ICPSR, a prominent social science data archive, and how ICPSR's data access and use controls have changed over a 50 year period in relationship to changes in scholarly practices, government regulation, technologies and sustainability concerns.

Sean Goggins

Goggins is an Assistant Professor at the University of Missouri Informatics Institute and School of Information Science and Learning Technologies. He teaches, publishes and conducts research on the uptake and use of information and communication technologies by small groups in medium to large-scale sociotechnical systems, from Facebook to online course systems. Goggins conceptualizes 'group informatics' as a methodological approach and ontology for making sense of the interactions between people in medium to large scale social computing environments. His research examines the information behavior, knowledge construction, identity development, performance and structural evolution of small, online groups. By understanding small groups of online learners, such as librarians or software engineers, and analyzing traces of their interactions, group informatics brings semantics to the clusters discovered in large scale social computing endeavors.

Libby Hemphill

Hemphill is an Assistant Professor in Communication and Information Studies at the Illinois Institute of Technology, where she also runs the Collaboration and Social Media Lab. Her work focuses on relationships among social media use, civic engagement, and diversity. She is especially interested in issues of access and how they relate to digital data – for instance how the technical barriers to collecting and managing digital data impact who uses such data in their research. She is a multi-methods researcher, and teaches courses on user research, social media, organizational communication, social network analysis and qualitative methods.

Steve Sawyer

Sawyer is Professor and Associate Dean for Research at the School of Information Studies at Syracuse University. He conducts multi-year, collaborative, research in the social informatics tradition -- with particular attention to the ways in which people organize to work together and use information and communication technologies. Sawyer leads courses that focus student's attention to the design, development and implementation of information systems, managing projects and systems, and to the roles of information and communication technologies relative to organizational and social change.

Kalpana Shankar

Shankar is a Lecturer in the School of Information and Library Studies at University College, Dublin. Her research focuses on the uses of data and data management practices (digital and otherwise) in the social and earth sciences, pervasive technology, and personal health. She draws from theory and practice in archival and information studies and STS (science, technology, and society), as well as related fields such as health informatics, computer supported cooperative work, and human-computer interaction.

Katie Shilton

Shilton is an Assistant Professor in the College of Information Studies at the University of Maryland. Her research examines the ethics and policy for the design of information technologies, systems and collections. She is currently working on a project investigating the ethical and social challenges raised by a new approach to internet architecture with the Named Data Networking initiative, as well as a project surveying

consumer expectations in the mobile data ecosystem. Shilton has published broadly on the intersection of values and design, including aspects of privacy and data.

References

- [1] Eschenfelder, K.R.; Johnson, A (forthcoming) "Governing the Data Commons: Controlled Sharing of Scholarly Data" *Journal of the American Society for Information Science and Technology*.
- [2] Goggins, S. P., Mascaro, C., & Valetto, G. (2013). Group informatics: A methodological approach and ontology for sociotechnical group research. *Journal of the American Society for Information Science and Technology*, 64(3), 516-539.
- [3] Hemphill, L., Roback, A. (2014) Tweet Acts: How Constituents Lobby Congress via Twitter. *Proceedings of the 2014 ACM Conference on Computer Supported Cooperative Work*, Baltimore, MD.
- [4] Kowalczyk, S., & Shankar, K. (2011). Data sharing in the sciences. *Annual review of information science and technology*, 45(1), 247-294.
- [5] Sawyer, S., Kaziunas, E., & Øesterlund, C. (2012). Social scientists and cyberinfrastructure: insights from a document perspective. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work* (pp. 931-934), ACM.
- [6] Shankar, K. (2012). Self-archiving and collaboration in science 2.0: an exploratory study. In *Proceedings of the 2012 iConference* (pp. 513-514), ACM.
- [7] Shilton, K. (2012). Participatory personal data: An emerging research challenge for the information sciences. *Journal of the American Society for Information Science and Technology*, 63(10), 1905-1915.