

Carol Tenopir, Professor Emerita, School of Information Sciences, University of Tennessee, Knoxville, Tennessee, USA ctenopir@utk.edu

Book review:

“We Exchange Data All the Time”: A case study on data conceptualization and data sharing by researchers in the context of open data policies

Madeleine Dutoit (2022.) Department of Archivistis, Library and Information Science, Faculty of Social Sciences, Oslo Metropolitan University

For the last decade and a half, studies of how researchers manage, describe, and share research data generated in the course of their work have been undertaken by researchers in many different countries and in a variety of subject disciplines. The previous studies often rely on anonymous surveys, either on a large multinational international scale (for example Tenopir et al. 2020) or a smaller regional or national scale or within a specific discipline (Tenopir et al. 2018). These surveys typically study current research data practices, willingness to share, and barriers to data sharing within the group that is being studied.

Results are extrapolated to the larger population, with the hope of predicting future data needs and practices. Sometimes in-depth interviews, such as those led by Borgman and her colleagues, are used to again focus on current practices and opinions that will help direct practice, while providing personal insights into individual behavior and future needs. For examples, see Borgman (2015). The nature of surveys and one-time interviews are that they provide self-reported data to measure the opinions of researchers at a single moment in time. Results rely on truthful answers and assume that the self-reports at the time of answering the survey reflect the sustained truth over time. Repeated surveys later usually do not study the same individuals, although they may study the same population. In studies over a decade, information scientists have found many barriers and motivations to data sharing. Barriers included lack of time, a perception of no reason to share, uncertainty over where to share, wanting to publish before sharing, and legal or privacy restrictions.

The main motivations to share and to study data sharing are the growing number of data sharing mandates (or promises of future mandates) from research funding agencies both in the European Union and the United States, and in many other countries around the world. The Open Science movement, which started with the push to make research articles available without charge to readers, is the underlying motivation as it expands to making research data also available to allow better understanding and replication of research and to guard against fraud or bad data collection practices. There are of course, some restrictions to making data open, such as proprietary data or data involving human subjects that is not anonymous.

Madeleine Dutoit's Doctoral Thesis builds on this ~15 year body of research, while providing new and deeper insights through her unique choice of methodology and theoretical underpinning. Dutoit centered her research on the broad topic of research data sharing, including how researchers conceptualize and define research data in the context of their work. She has chosen the methodological approach of a Case Study, focusing on an international multidisciplinary research team to study their data conceptualization and opinions of data and practices of data sharing in the context of a large-scale funded research project. The case study focuses on a distributed interdisciplinary scientific research group funded by Horizon 2020 in the EU for a five-year project, of which she observed year two.

Wenger's community of practice forms the conceptual theoretical framework for this thesis, within a broader group of theories surrounding practice theories (Wenger, 1998). Shared practices of what

participants do and create within a shared task creates a community of practice, which in this study allows the case study results of what the community believes and does regarding data practices over time to be fitted into a broader theoretical underpinning.

Eighteen researchers within the international interdisciplinary research team were interviewed and observed over a year and documents and data they created were analyzed. This variety of methods within the case study approach allows the author to discover insights into how various members of the interdisciplinary research group think of and define data at multiple stages in the project, what their data look like, what they think about data sharing, and what they actually do. By studying one community for an entire year, the researcher observed how those views and plans may change over the course of working on an extended research project.

Dutoit uncovered the complexities of how a distributed research team approach research data. The researchers did not all agree even on what data is (or are) and, although they formulated a data management plan together at the outset of the project as required by the funder, they did not all agree overtime on the usefulness of the plan or the practicalities of managing or sharing their data according to the plan. Creating an agreed-upon data management plan and subsequently following it or agreeing how to follow it were revealed to be in conflict and there were a variety of conflicting reactions expressed about the need or utility of data management plans or, specifically, the ease of using the DMP tool available to them to meet the funders' requirements.

Dutoit observed ways a research community of practice was built relating to data and how these communities changed over time, as well as the complex nature of data, data policies, and interactions. She found major differences between theoreticians and experimentalists and a disconnect between the data management plan and behaviors over time. The findings and insights of this thesis will benefit the actors responsible for data management planning, which include funders, librarians, research group managers, and the researchers themselves.

The strengths of this thesis are many, including a rigorous use of the long-term (one year) ethnographic case study approach, something that has rarely been used in studies of data sharing. This has allowed her to reveal the complex nature of how researchers conceive data in their work and how that may change depending not only on the type of data they generate, but on the methods they

use, the stage of their research, and their interactions with others in the group regarding data. Dutoit found that even the definition of data can be fluid among researchers, an interesting finding in itself. Her insights into differences between experimentalists and theoreticians provides insights for the information science research community and others beyond the usual focus just on subject discipline differences.

In the thesis, Dutoit herself acknowledges the main weakness of the case study approach, in that results may not be generalizable, although she does provide examples of how her findings may be typical of similar research projects. Her one-year time span allowed deep exploration, but it did not allow her to see the full length of the project. I suspect that few new insights would have been revealed with a longer term, but it is possible. She revealed the disciplines of only two (outlier) participants, so disciplinary differences cannot be expanded on.

In summary, mandates for open data often do not appreciate the complexities that Dutoit has uncovered and, therefore, these results should be of great interest to a variety of stakeholders. The disconnect regarding data management between researchers and policy makers is uncovered in this case study, yet it concludes with recommendations that will be helpful to making research data management mandates more successful in the future.

References

Tenopir, C., Rice, N., Allard, S., Baird, L., Borycz, J., Christian, L., Frame, M., Grant, B., Olendorf, R., Sandusky, R., & Zolly, L. (2020). 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, C., Christian, L., Allard, S., & Borycz, J. (2018). Research data sharing: Practices and attitudes of geophysicists. *Earth and Space Science*, 5(12), 891-902. <https://doi.org/10.1029/2018ea000461>

Borgman, C. (2015). *Big data, little data, no data: Scholarship in the networked world*. Cambridge, MA: MIT Press.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity*. Cambridge: Cambridge University Press.