

Conceptualizing Data Needs within Contexts of Data Discoverability and Reuse: A Study of Environmental and Social Scientists

Authors: Ying-Hsang Liu¹, Isto Huvila¹, Jessica Kaiser¹, Zanna Friberg¹, Olle Sköld¹, Lisa Andersson¹, Megan Power², Mingfang Wu³

Author's Affiliations: ¹ Uppsala University, ² Monash University, ³ Australian Research Data Commons

Keywords: Data needs, Data discovery contexts, Data reuse, Environmental scientists, Social scientists

STRUCTURED ABSTRACT

Aim of your contribution

This study contributes to the conference theme of processes and practices in information science by investigating the research data discovery contexts of ecological and social scientists who are engaged with reusing datasets for research purposes.

Value of your contribution

Our study aimed to gain insight into the data discovery contexts of ecological and social scientists and to understand the data needs related to data discoverability and reuse. Our research identified four dimensions of data needs, including 1) Research processes, 2) Making sense of data, 3) Data reuse, and 4) Data access. Additionally, we have proposed a conceptualization of data needs within the context of data reuse, which has not been thoroughly examined in previous studies.

Research outline

- Research context

The open data movement has witnessed significant expansion in recent years, leading to an increase in the number of stakeholders, such as research institutions, funding agencies, data practitioners, data service providers, and end-users, showing a keen interest in exploring and reusing data. In Australia, the Australian Research Data Commons (ARDC) has taken the lead role in supporting the development of collaborative research data systems and services. This project aims to examine the different contexts that ecological and social scientists discover data to examine how these contexts influence data discoverability and facilitate data reuse.

- Problem/overall research aims/research question

As part of a larger research project, this presentation focuses on characterizing data needs and the contexts of data reuse to inform the design and evaluation of data discovery systems and services. Our research questions are:

1. How do we conceptualize data needs from user perspectives?
2. What is the relationship between the data needs and the contexts of data reuse?

- Design/methodology/ research approach

The aim of this study is to identify the different contexts in which data is discovered, using a mixed-method approach within the post-positivist research paradigm (Williamson, 2018). To achieve this goal, we employed a combination of survey and in-depth interview techniques to investigate the broader contexts of data discovery in people's information-seeking processes. A pre-interview survey was conducted to gather information about the participants' background, including their research areas/topics, career stages, job roles, and data sources used in recent projects. This information was utilized in the subsequent semi-structured in-depth interview. To elicit the contexts of data discovery, we used a critical incident technique, which involved asking probing questions (Davenport, 2010; Flanagan, 1954). The interview protocol was structured based on the stages of a data lifecycle adapted from Yong et al. (2021).

We had the opportunity to conduct interviews with 24 participants in partnership with three organizations: TERN (Australia's Terrestrial Ecosystem Research Network), ADA (The Australian Data Archive) and CSIRO (The Commonwealth Scientific and Industrial Research Organisation). The participants held diverse job roles and were at different stages of their career. While most of them were involved in collaborative research projects and had extensive experience in discovering data, we also interviewed PhD candidates and academic professionals who focused on teaching, bringing varied perspectives to the study. Additionally, we were able to interview a few researchers who had transitioned into data practitioners (all from our partners), and we observed that they were more deeply involved in the various stages of a data lifecycle. We used the nVivo qualitative data analysis tool to analyze the interview transcripts, using the grounded theory approach (Strauss & Corbin, 2014)

- Findings

Our study has identified four dimensions of data needs, namely 1) Research processes, 2) Making sense of data, 3) Data reuse, and 4) Data access, and examined their relationship with the roles of data managers and end-users. Our findings contribute to the existing literature on data needs, building upon recent studies that have explored data reuse (Thomer, 2022), researchers' requirements for sharing data across different career stages (Stahlman, 2022), and the conceptualization of data needs within local and dynamic contexts (Gregory, 2022). Our research also emphasizes the potential usefulness of research data (Huvila, 2022) and the need for paradata (Börjesson et al., 2022).

- Research limitations/implications

Our findings suggest that anticipating the contexts of data reuse involves considering what data users may find useful. Ensuring data quality is crucial for

successful data reuse, which involves having access to organizational data expertise, providing data in various formats and platforms, and ensuring sufficient data coverage. Data needs are influenced by the specific research objectives, which in turn affects the criteria for selecting and reusing data. Clear data licensing conditions are crucial to facilitate data reuse. Transferability of the findings may be limited by the context dependent and dynamic nature of data needs in different settings.

References

- Börjesson, L., Huvila, I., & Sköld, O. (2022). Information needs on research data creation. *Information Research*, 27(Special Issue), isic2208. <https://doi.org/10.47989/irisic2208>
- Davenport, E. (2010). Confessional methods and everyday life information seeking. *Annual Review of Information Science and Technology*, 44, 533–562. <https://doi.org/10.1002/aris.2010.1440440119>
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4): 327–359. <https://doi.org/10.1037/h0061470>
- Huvila, I. (2022). Improving the usefulness of research data with better paradata. *Open Information Science*, 6(1), 28–48. <https://doi.org/10.1515/opis-2022-0129>
- Gregory, K., & Koesten, L. (2022). *Human-centered data discovery*. Springer International Publishing. https://doi.org/10.1007/978-3-031-18223-5_3
- Stahlman, G. R. (2022). From nostalgia to knowledge: Considering the personal dimensions of data lifecycles. *Journal of the Association for Information Science and Technology*, 73(12), 1692–1705. <https://doi.org/10.1002/asi.24687>
- Strauss, A., & Corbin, J. (2014). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (4th ed.). SAGE.
- Thomer, A. K. (2022). Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits. *Journal of the Association for Information Science and Technology*, 73(8), 1155–1170. <https://doi.org/10.1002/asi.24620>
- Williamson, K. (2018). Research concepts. In K. Williamson & G. (Eds.), *Research Methods: Information, Systems, and Contexts* (2nd ed., pp. 3–25). Chandos Publishing. <https://doi.org/10.1016/B978-0-08-102220-7.00001-7>
- Yong, A., Zahuranec, A. and Verhulst, S. (2021). A layered approach to documenting how the third wave of open data can provide societal value. Accessed on March 7, 2023 at: <https://opendatapolicylab.org/articles/the-onion-model-a-layered-approach-to-documenting-how-the-third-wave-of-open-data-can-provide-societal-value/index.html>

Suggested Citation

Liu, Y.-H., Huvila, I., Kaiser, J., Friberg, Z., Sköld, O., Andersson, L., Power, M., & Wu, M. (2023). Conceptualizing data needs within contexts of data reuse: A study of ecological and social scientists. *Trends in Information Science – The ASIS&ST European Chapter Research Series*. June 20-21, 2023. Virtual conference.