INFORMATION STUDIES DAYS 2024

Paradata literacy and the challenges of research data management

Isto Huvila
Uppsala University
isto.huvila@abm.uu.se
https://orcid.org/0000-0001-9196-2106

Jessica Kaiser
Uppsala University
jessica.kaiser@abm.uu.se
https://orcid.org/0000-0002-4308-8888

Olle Sköld
Uppsala University
olle.skold@abm.uu.se
https://orcid.org/0000-0002-0904-7222

Lisa Andersson
Uppsala University
Karolinska Institutet
lisa.andersson@ki.se
https://orcid.org/0000-0003-0493-6845

Keywords: information, literacy, processes, research data

Pysyvä osoite: https://doi.org/10.23978/inf.148594

Introduction

Recent research data management literature has increasingly underlined the importance of broadening the scope of documenting preserved datasets to encompass contextual information (Faniel et al., 2019) and paradata, i.e. data on the processes of its inception and processing (Huvila, 2022). It has been further noted that this has an impact on the competencies required from data creators, managers, and users.

The aim of this paper is to provide insights into what types of competencies are necessary for successful data documentation and reuse with a focus on the literacies pertaining to paradata, or documentation of data practices and processes. We draw from an international interview study of researchers and professionals (N=33) working with archaeological data to delve into the question of the elements of paradata literacy, what competencies are desired of whom and for what purposes, and how paradata literacy or literacies link to the broader genre of data literacies in the context of data-work.

Paradata and paradata literacy

A rapidly growing corpus of research addresses how to adequately document data creation, processing, and use to support research data reuse. The work links to broader interdisciplinary endeavours of inquiring into paradata as a means to document processes and practices (Cameron, Franks & Hamidzadeh, 2023). The studies show that information qualifying as paradata can be found in different parts of research and data documentation in various forms and formats (e.g., Huvila & Sinnamon, 2022; Huvila et al., 2021; Börjesson et al., 2022) and that even if they align in principle, data creators' and reusers' paradata needs differ from each other (e.g., Börjesson et al., 2022; Huvila et al., 2024).

So far there is scant research specifically on paradata-related competencies. Several researchers have, however, pointed out the importance of including structural (Kansa & Kansa, 2018), contextual, and processual expertise as a part of necessary data-related competences. Even if data is formally findable, accessible, interoperable, and reusable according to the widely adopted FAIR principles, specific competencies are required to assess its adequacy for purpose and use (Bishop et al., 2019).

Methods and material

We conducted semi-structured interviews (A-AG) online with 33 archaeologists and cultural heritage professionals. A purposive theoretical sampling was used to ensure as much breadth as possible with regard to engagement with research data, subject specialisations and career stages. The interviewees were based in multiple countries around the world, represented different genders, and had a mix of experience in data creation, reuse, and management. An interview guide with questions on data creation and reuse developed based on earlier research on researchers' needs for contextual information was used. Particular focus was placed on the creation and use of paradata about research data. The interviews were transcribed by a professional transcriber, and analysed by close reading to identify passages relating to data-related skills, competencies and literacies.

Findings

Some interviewees entertained the idea of generic data competencies as a distinct complement to domain expertise (e.g., C, K, O) while others saw little value in generic training in data-making (e.g., P, AD). The first group included those with data management experience whereas the latter was typical for data creators and users. However, the critical competence emphasised across the interview record was archaeological expertise, and, more specifically, literacy within specific subdomains of archaeology (P, AD) that incorporate a working knowledge of field-specific practices and vocabulary (AD). For a colleague with working domain knowledge, a simple written description could be enough for communicating critical paradata (O) whereas for someone from outside of the domain, making data intelligible could be close to impossible (P). A basic knowledge of working with data and databases was seen as a complement to domain expertise but as a facilitator rather than a kernel of what qualifies as (para)data literacy (e.g., A, M, P). At the same time, determining what qualifies as a sufficient level of competence was considered difficult and contextual (P, Z).

Discussion and conclusions

The findings suggest firstly that the key to successful paradata literacy lies in a robust understanding of context-specific data practices, rather than in the mastery of highly advanced,

general data-making skills. Domain specialists focused on domain-specific competencies, as an underpinning of both successful data creation and use, whereas data managers tended to emphasise, at least to some degree, the relevance of more generic skills. The findings align with earlier observations of the variation and *ad hoc* nature of archaeologists' data literacy (Kaiser, forthcoming) pointing to a need for more systematic consideration of its constituents across the discipline.

Secondly, the findings further indicate a paradigmatic breach between domain-oriented and data-focused data-work and their respective literacies. This gap is often framed in the popular deficiency discourse as an inadequacy in researchers' data literacy. However, if the obvious sufficiency of domain-specific data practices is accepted, data literacy can also be seen as a relational undertaking, enacted through collaborative data-work of data creators, managers and users (cf. Friberg & Huvila, forthcoming). Instead of being conceived of as a generic competence applied to diverse disciplines, data literacy should perhaps be framed as a family of, to various degrees and with varying affinities to each other, interlinked proficient (sub)disciplinary practices of dealing with data and understanding its workings (cf. Börjesson et al., 2022). As Burton et al. underline, acknowledging that data is not given entails a shift from datasets to data settings and data literacies the literacies of data infrastructures (Burton, 2023) Rather than focusing solely on mastering a limited set of practices, as Sander (2024) suggests in the context of critical datafication literacy, (para)data literacy should also empower individuals to take different situationally relevant forms of action.

Finally, the findings highlight the difficulty in making a watertight distinction between paradata- and data literacy. However, as the data competencies discussed by the interviewees were predominantly relating not to the data itself but rather to the context and practices in which it was generated, there is a reason to argue that data literacy is about mastering paradata — paradata literacy — rather than attaining competence in the data itself.

References

Bishop, B. W., Hank, C., Webster, J., & Howard, R. (2019). Scientists' data discovery and reuse behavior: (Meta)data fitness for use and the FAIR data principles. *Proceedings of the Association for Information Science and Technology, 56*(1), 21–31. https://doi.org/10.1002/pra2.4

Börjesson, L., Huvila, I., & Sköld, O. (2022). Information Needs on Research Data Creation. Information Research, 27(Special Issue). https://doi.org/10.47989/irisic2208

Börjesson, L., Sköld, O., Friberg, Z., Löwenborg, D., Pálsson, G., & Huvila, I. (2022). Re-Purposing Excavation Database Content As Paradata: An Explorative Analysis of Paradata Identification Challenges and Opportunities. *KULA: Knowledge Creation, Dissemination, and Preservation Studies, 6*(3), 1–18. https://doi.org/10.18357/kula.221

Burton, A. G., & Chun, W. H. K. (2023). *Algorithmic Authenticity: An Overview*. Meson press.

Cameron, S., Franks, P. & Hamidzadeh, B. (2023). Positioning Paradata: A Conceptual Frame for AI Processual Documentation in Archives and Recordkeeping Contexts. *Journal on Computing and Cultural Heritage, 16*(4), 1–19. https://doi.org/10.1145/3594728

Faniel, I. M., Frank, R. D. & Yakel, E. (2019). Context from the Data Reuser's Point of View. *Journal of Documentation, 75*(6), 1274–1297. https://doi.org/10.1108/JD-08-2018-0133

Friberg, Z., & Huvila, I. (forthcoming). Methods for Managing Paradata. In I. Huvila (Ed.), *Paradata: Documenting Data Creation, Curation and Use*.

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

Huvila, I., Andersson, L., & Sköld, O. (2024). Patterns in Paradata Preferences among the Makers and Reusers of Archaeological Data. Data and Information Management. https://doi.org/10.1016/j.dim.2024.100077

Huvila, I., & Sinnamon, L. (2022). Sharing Research Design, Methods and Process Information in and out of Academia. Proceedings of the Association for Information Science and Technology, 59(1), 132–144. https://doi.org/10.1002/pra2.611

Huvila, I., Sköld, O. & Börjesson, L. (2021). Documenting Information Making in Archaeological Field Reports. *Journal of Documentation, 77*(5), 1107–1127. https://www.emerald.com/insight/content/doi/10.1108/JD-11-2020-0188/full/html

Kaiser, J. (forthcoming). On Babies and Bathwater: Strategies for Dealing with Imperfect Data in (Bio)Archaeology.

Kansa, S. W., & Kansa, E. C. (2018). Data Beyond the Archive in Digital Archaeology: An Introduction to the Special Section. *Advances in Archaeological Practice, 6*(2), 89–92. https://doi.org/10.1017/aap.2018.7

Sander, I. (2024). *Critical Datafication Literacy: A Framework and Practical Approaches*. 1. Auflage. Digitalität Und Bildung 2. Bielefeld: transcript.