Co-Designing AI Assistants for Research Data Management Through Interactive Support Sessions

Eva Eleonora Ferradosa [1], Christoph Lofi [1]

[1] Department of Software Technology, Delft University of Technology, The Netherlands
{e.e.ferradosa,c.lofi}@tudelft.nl

Metadata documentation keeps datasets findable, interpretable, and reusable [5]. Yet for many researchers it still feels like a chore: form-based documentation is time-consuming and effortful [4], and rigid forms do not reflect the way researchers naturally share context about their experiments, which is conversational, descriptive, and rich in tacit knowledge [3]. So far, metadata management has largely been treated as a technical or compliance problem: yet another standard or guide is handed to individual researchers, who are expected to figure it out on their own [2]. Data stewards can help, but many projects lack dedicated stewardship, and stewards' time does not scale with growing demand.

Our work explores agentic (AI-based) assistants for Research Data Management (RDM) that not only support metadata creation, but are also aligned with the way researchers naturally communicate. Significant research is still required to determine how to make such an assistant genuinely helpful in real-world workflows: which tasks it should take on, and how it should interact with researchers. Our guiding research question is therefore: *what should an AI assistant actually be able to do to be genuinely useful in real RDM workflows?*

To answer this, we are starting a study based on human-led, interactive RDM support sessions. In short (30–60 min.) sessions, researchers receive hands-on help documenting their (meta)data. Besides giving participants immediate support, the sessions produce conversational transcripts and insights, which we analyze to surface real needs, friction points, and the solutions negotiated between facilitator and researcher. The sessions start from an unstructured format, and their structure will be refined iteratively as insights emerge from the sessions themselves. The study is situated in CropXR, a Dutch institute developing more resilient crops through data-driven plant research [1]. At DBDBD, we will present the motivation and design of this ongoing study and welcome discussion with the community. Ultimately, we aim for future RDM tools to be shaped by practice: fitting the tool to the researcher, not the other way around.

References

[1] CropXR: Making crops more resilient, sustainable, and climate-adapted. https://cropxr.org/, 2026.

[2] Birkbeck, G., Nagle, T., and Sammon, D. Challenges in research data management practices: a literature analysis. *Journal of Decision Systems 31*, sup1 (2022), 153–167.

[3] Mayernik, M. S. Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science 49*, 5 (2019), 732–757.

[4] Perrier, L., Blondal, E., and MacDonald, H. The views, perspectives, and experiences of academic researchers with data sharing and reuse: A meta-synthesis. *PLoS ONE 15*, 2 (2020), e0229182.

[5] Wilkinson, M. D., et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data 3*, 1 (Mar. 2016).