



Delft University of Technology

Document Version

Accepted author manuscript

Licence

Dutch Copyright Act (Article 25fa)

Citation (APA)

Ferradosa, E. E., & Lofi, C. (2026). *Collaborative Metadata Documentation as a Sociotechnical Practice: Lessons from a Data Harvest Day in a Plant Systems Biology Research Consortium*. Abstract from NFDI4LifeScience.

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Collaborative Metadata Documentation as a Sociotechnical Practice: Lessons from a Data Harvest Day in a Plant Systems Biology Research Consortium

Eva Eleonora Ferradosa, Christoph Lofi
TU Delft, The Netherlands

Introduction

In large-scale research consortia involving collaborations between academia and industry, the ability to find, reuse, and integrate experimental data depends critically on the availability of high-quality, standardised metadata. In such settings, experimental researchers are increasingly expected to document their data in accordance with community standards before publication, as a prerequisite for internal collaboration and long-term reuse. However, metadata documentation is routinely relegated to the bottom of researchers' priority lists not only because it is a labour-intensive and cognitively demanding activity, but also because its purpose and value are unclear to researchers compared to the effort it requires. Metadata documentation calls for familiarisation with standards through extensive reading of often difficult-to-interpret material, mapping complex experimental designs onto subjective metadata models, and the repetitive manual entry of metadata values. Since researchers perceive future data reuse as uncertain, pre-emptive metadata documentation is often seen as not being worth the trouble. This leads to standard metadata being frequently forgotten, left incomplete, or inconsistent, limiting the potential for data-driven collaboration in transdisciplinary research environments.

Context

Within the CropXR consortium, a transdisciplinary research initiative focused on developing climate-resilient crops through data-driven plant systems biology, the mentioned challenges became particularly salient. CropXR integrates heterogeneous data types, including phenotyping and sequencing data, and relies on metadata standards drawn and adapted from the broader life sciences community. In practice, metadata documentation in the consortium is structured around a conceptual model inspired primarily by the ISA (Investigation–Study–Assay) framework, with more specific elements informed by standards such as MIAPPE and ENA. Filling out the necessary metadata forms for a Study is estimated to take as long as eight hours. In response to these challenges, the CropXR consortium organised a one-day in-person “*Data Harvest Day*”, motivated by the hypothesis that metadata documentation could be approached more effectively as a shared, collaborative activity rather than an individual compliance task carried out in isolation. The intention was not only to provide technical assistance but also to create a collective setting in which researchers could discuss, interpret, and make sense of metadata standards together, thereby addressing both conceptual and motivational barriers. Furthermore, since concerns were expressed regarding ownership, confidentiality, and access within a complex governance structure, these topics were addressed and clarified, as they can negatively influence engagement in metadata documentation.

This contribution reports lessons learned from this intervention. The analysis draws on three empirical sources: (1) observational notes taken during the event, (2) an anonymous post-event questionnaire, and (3) reflective discussions among the organisers. Particular attention is given to how collaborative documentation affected participants' understanding, confidence, motivation, perceived productivity, and sense of burden, as well as to qualitative insights into why these effects occurred.

Approach

Twenty experimental researchers at different stages of their projects participated voluntarily in the event. They were supported by five members of the consortium's data team, who, while not formally designated as data stewards, collectively acted in facilitative roles during the event. An introductory session outlined the importance of metadata documentation for data collaboration within the consortium. It explicitly addressed concerns regarding data confidentiality and access, emphasising that cataloguing data did not equate to unrestricted sharing. This was followed by tutorial-style explanations of existing procedures, including how to upload data to a restricted-access research drive via a web interface or command-line tools, how to register new studies, and how the consortium's metadata model is structured at a conceptual level. The majority of the day was then dedicated to hands-on work, during which participants were encouraged to begin documenting their metadata while receiving one-to-one and small-group support, and to openly discuss questions, uncertainties, and interpretations of the standards in use. At the end of the

event, participants voluntarily completed a questionnaire containing Likert, multiple-choice, and open-ended questions about collaborative metadata documentation, and a final feedback discussion was held.

Quantitative Results

Quantitative questionnaire results, collected from 11 of 20 participants, indicate that collaborating on metadata documentation had a strongly positive effect on participants' understanding of what metadata needs to be documented: 90.9% of respondents reported a positive effect and 9.1% a very positive effect. Similarly, 81.8% reported a positive (72.7%) or very positive (9.1%) impact on their confidence in documenting metadata, with 18.2% reporting no effect, and 72.7% reported a positive (54.5%) or very positive (18.2%) impact on motivation, with 27.3% reporting no effect.

By contrast, perceived productivity and burden showed more heterogeneous outcomes. While 54.5% of respondents reported increased productivity when working collaboratively, 36.4% observed no effect, and 9.1% experienced a negative effect. Likewise, although 18.2% reported that collaboration reduced the perceived burden of metadata documentation, 45.5% experienced an increased sense of burden, and 36.4% reported no change.

Importantly, despite these mixed perceptions, all respondents expressed willingness to participate in a similar session again, with more than 80% indicating moderate to extreme willingness, and preferences emerging for quarterly, half-day, in-person sessions (preferred by 50%, 82%, and 91% of respondents, respectively).

Qualitative Insights and Interpretation

Qualitative questionnaire responses and observational notes provide essential context for interpreting the quantitative findings. Rather than technical incapacity at the tool level, the most persistent difficulties related to conceptual understanding of metadata standards. Participants struggled to align and scope the abstract entities defined by the ISA framework, such as investigations, studies, and assays, with the realities of their experimental workflows. Questions frequently arose regarding what constituted relevant metadata, how much detail was sufficient, and at what stage of the research lifecycle metadata should be documented. These challenges were often compounded by uncertainty about the purpose of metadata documentation, particularly in relation to future reuse scenarios that were not yet clearly defined. Debate and negotiation helped alleviate these uncertainties and shed light on how to proceed in different cases.

At the confidence and motivational level, working collaboratively was widely appreciated. Participants valued the opportunity to discuss metadata together, to hear how others interpreted similar requirements, and to receive individualised feedback tailored to their specific experimental situations. Several respondents emphasised that every experiment seemed “special” and difficult to standardise without discussion, and observational notes highlighted that being understood and taken seriously was often more important to participants than receiving prescriptive instructions. The role of the supporting data team emerged as particularly significant. Participants valued the presence of interdisciplinary facilitators who could help interpret standards, receive feedback, troubleshoot access issues, and address governance-related questions in real time. Although formal data stewardship roles were absent, stewardship was enacted in practice through availability, credibility, and empathy. The social aspect of the session had, therefore, a palpable effect on the confidence and motivation of researchers in tackling the daunting task of metadata documentation. Requests to include additional expertise, such as bioinformaticians or modellers, further suggest that researchers see metadata documentation as situated at the interface between experimental practice, analysis, and infrastructure, rather than as a purely administrative task.

At the same time, group discussions also risked increasing confusion for some participants, particularly when plenary sessions mixed general explanations with highly specific questions. This tension underscores the importance of carefully balancing collective discussion with individualised support in the design of such interventions. The mixed perceptions of productivity and burden indicate that collaboration can temporarily surface complexity more clearly, making the work feel heavier for some, even as it resolves conceptual confusion and lays the foundation for more sustainable practices.

Governance and trust considerations formed a critical backdrop to metadata discussions. Questions about data ownership, confidentiality, responsibilities, and access rights were frequently raised, reflecting the complexity of a consortium involving multiple funding schemes and project structures. Questionnaire feedback highlighted the importance of explicitly communicating that the research drive and metadata catalogue were intended to support collaboration rather than impose open data requirements, with creators retaining control over visibility and access, and that uncertainty in this area negatively affected researchers' willingness to engage with metadata documentation.

Conclusion

Taken together, these results suggest that collaborative metadata documentation functions less as a mechanism to immediately increase the quantity of documented metadata and more as a sociotechnical intervention that improves shared understanding, confidence, and motivation. From a research data management perspective, the findings imply that metadata documentation should be treated primarily as a sense-making activity supported by dialogue, rather than as a purely technical task driven by compliance expectations. Early-stage interventions should prioritise understanding and alignment over completeness, recognise stewardship as a distributed practice rather than a formal role, design tools and templates that guide interpretation as well as metadata entry, and address governance and trust explicitly as integral components of metadata training. In this way, collaborative formats such as the “*Data Harvest Day*” can help foster a data-sharing culture that is understood, collectively owned, and more likely to endure in complex, data-intensive research environments.