



Delft University of Technology

Document Version

Accepted author manuscript

Citation (APA)

Ferradosa, E. E., & Lofi, C. (2026). *Rethinking Metadata Documentation: From Technical Solutions to Sociotechnical Support for Sustainable Research Data Management*. Poster session presented at Association for Computing Machinery - womENCourage™ 2026.

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.

Rethinking Metadata Documentation

From Technical Solutions to Sociotechnical Support for Sustainable Research Data Management

Eva Eleonora Ferradosa
Delft University of Technology
Delft, The Netherlands
e.e.ferradosa@tudelft.nl

Christoph Lofi
Delft University of Technology
Delft, The Netherlands
c.lofi@tudelft.nl

ACM Reference Format:

Eva Eleonora Ferradosa and Christoph Lofi. 2026. Rethinking Metadata Documentation: From Technical Solutions to Sociotechnical Support for Sustainable Research Data Management. In *Proceedings of 13th ACM Celebration of Women in Computing (WomENCourage'26)*. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 Introduction

Research metadata refers to structured and contextual information describing how datasets are created, by whom, under which conditions, and how they can be accessed and interpreted [5]. Beyond supporting findability, metadata captures assumptions and decisions not explicit in the data, enabling long-term reuse, integration, and responsible data-driven research [2]. Over the past decade, extensive effort has been invested in defining metadata standards [4], building data platforms [7], and promoting principles such as FAIR [6] to improve metadata quality and reusability. Despite this technical progress, metadata documentation remains a persistent challenge in research practice. Researchers widely acknowledge the importance of metadata documentation, particularly when required for data publication or journal submission; however, it is often perceived as a repetitive, time-consuming, and cognitively demanding task that competes with other high-attention research activities [1]. In such mandatory contexts, researchers typically comply with metadata requirements, but often with minimal effort or care, as documentation is treated as an obligation rather than as a meaningful research practice.

In this work, we discuss why current approaches to metadata documentation remain overly focused on technical solutions, while underestimating the sociotechnical conditions under which metadata is produced in practice. We present a research agenda that reframes metadata documentation as a sociotechnical practice shaped by researcher motivation, shared understanding, and community support. By drawing on experiences from an interdisciplinary research consortium, this work explores how collaborative documentation practices, stewardship, and emerging tooling can better support researcher wellbeing while strengthening the long-term sustainability of research data infrastructures.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
WomENCourage'26, Nice, France
© 2026 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-XXXX-X/2018/06
<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

2 The Traditional Approach to Metadata

In everyday research workflows, metadata documentation is frequently minimized or postponed. Researchers face limited time and increasing responsibilities, alongside multiple metadata standards, platform-specific requirements, and extensive documentation. Metadata is often produced primarily to satisfy publication requirements rather than as a meaningful part of research practice. As a result, researchers often rely on reactive, informal communication with data users to convey contextual information [3], with details about data structure, assumptions, and limitations shared through emails, messages, or meetings rather than through metadata records. From the researchers' perspective, this approach is rational: documenting metadata comprehensively requires substantial effort, while the likelihood of future reuse is uncertain. However, this practice has significant long-term consequences. Contextual knowledge shared informally is rarely preserved, and when reuse is attempted years later, original researchers may no longer recall key details or be accessible. In such cases, formally stored data can become practically unusable, undermining the goals metadata systems are intended to achieve.

Most existing research on metadata management addresses these challenges through technical solutions, focusing on metadata standards, ontologies, and platforms intended to support pre-emptive documentation. These solutions are typically presented to researchers as resources they should adopt, with the implicit expectation that improved tooling will translate into improved metadata. In practice, responsibility for deciding what to document, to what extent, and when is largely placed on individual researchers, who are expected to anticipate future reuse, interpret abstract standards, and choose between numerous tools and guidelines. When metadata proves insufficient, accountability is often attributed to individual decisions, despite the inherent uncertainty of predicting future reuse. Data stewards may provide guidance, but often face similar uncertainties and therefore point researchers to existing standards and resources, reinforcing a sense of being overwhelmed rather than supported in decision-making. Consequently, researchers are left navigating a complex sociotechnical landscape largely on their own.

3 Metadata as a Sociotechnical Practice

Metadata documentation requires substantial cognitive effort, including reflection on experimental design, articulation of tacit knowledge, and interpretation of standards. When this effort is invisible, unsupported, or unrecognized, it can become demotivating. Sustained engagement with metadata work depends not only on tools, but also on social support, shared understanding, and recognition of effort within the research community. From this perspective, effective metadata practices require collective negotiation.

Researchers benefit from discussing with peers what information is most relevant, what can reasonably be documented, and how standards should be interpreted in real-world contexts. Motivation to invest effort increases when documentation choices are informed by shared expectations and when the value of that effort is visible. For metadata standards to function effectively, data creators and data users need a shared understanding of what is prioritized and why. When documentation requirements remain abstract, metadata reflects imagined reuse needs rather than articulated ones.

Bringing data creators and data users into conversation creates opportunities to align documentation practices with actual reuse interests, reducing uncertainty and unnecessary effort. Data stewards can play a key role in facilitating these exchanges. Beyond technical guidance, they can support discussion, mediate between standards and practice, and help translate reuse needs into actionable documentation strategies. To do so effectively, data stewards must be embedded within the research community and informed by ongoing dialogue between data producers and data users. Creating space for such dialogue not only clarifies documentation priorities, but also reduces isolation and uncertainty, supporting more sustainable and humane ways of working with data-intensive systems.

4 Rethinking Metadata Work

To support researchers working under increasingly data-driven environments, we investigate metadata documentation as a sociotechnical practice rather than a purely technical or compliance-oriented task. Conducted as a PhD project within the CropXR consortium, an interdisciplinary public–private initiative focused on data-driven crop resilience research, this work examines how researcher motivation, shared understanding, and community support shape metadata practices in real-world settings.

Within the CropXR consortium, these sociotechnical principles are explored through practices such as Data Harvest Days, where researchers document metadata collaboratively during dedicated sessions facilitated by data stewards. These sessions provide space to articulate uncertainties, discuss how standards fit specific datasets, and negotiate documentation priorities collectively. Planned future iterations aim to include data users alongside data creators, further aligning documentation practices with reuse needs. In parallel, this research investigates how AI-based tools can support sociotechnical metadata practices by reducing manual and cognitive effort. Automated metadata extraction from existing research artifacts, such as electronic lab notebooks, documentation and logs, or transcripts from data discussions and interviews, can capture information in ways that align with how researchers naturally communicate, without replacing human judgement.

The overarching goal of this research is to improve the quality and long-term sustainability of research metadata. Progress toward this goal is explored through empirical observations of researcher engagement, perceived effort, and the richness of contextual information captured during collaborative documentation activities. These emerging practices reflect a shift toward sociotechnical systems that value researcher voice, shared responsibility, and long-term engagement, aligning metadata work with broader goals of wellbeing and sustainability in computing.

5 Conclusion and Outlook

Metadata documentation is essential for sustainable research data infrastructures, but it cannot remain an isolated and burdensome task carried out by researchers alone. Improving metadata quality requires sociotechnical systems that recognize cognitive effort, support collective decision-making, and value the human work involved in documentation. By reframing metadata documentation as a shared and supported practice, this research contributes to broader discussions on wellbeing in computing, while arguing that community-based support for documentation work is foundational for efficient data reuse and long-term scientific progress.

Acknowledgments

The author gratefully acknowledges the collaboration of the members of the CropXR consortium, whose openness to discussion and experimentation has been essential to this work.

References

- [1] Dylan G. E. Gomes, Patrice Pottier, Robert Crystal-Ornelas, Emma J. Hudgins, Vivienne Foroughirad, Luna L. Sánchez-Reyes, Rachel Turba, Paula Andrea Martinez, David Moreau, Michael G. Bertram, Cooper A. Smout, and Kaitlyn M. Gaynor. 2022. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *Proceedings of the Royal Society B: Biological Sciences* 289, 1987 (Nov. 2022). doi:10.1098/rspb.2022.1113
- [2] Jeremy Leipzig, Daniel Nüst, Charles Tapley Hoyt, Karthik Ram, and Jane Greenberg. 2021. The role of metadata in reproducible computational research. *Patterns* 2, 9 (2021), 100322. doi:10.1016/j.patter.2021.100322
- [3] Matthew S Mayernik. 2019. Metadata accounts: Achieving data and evidence in scientific research. *Social Studies of Science* 49, 5 (2019), 732–757. arXiv:https://doi.org/10.1177/0306312719863494 doi:10.1177/0306312719863494 PMID: 31354073.
- [4] Susanna-Assunta Sansone, Philippe Rocca-Serra, Dawn Field, Eamonn Maguire, Chris Taylor, Oliver Hofmann, Hong Fang, Steffen Neumann, Weida Tong, Linda Amaral-Zettler, Kimberly Begley, Tim Booth, Lydie Bougueleret, Gully Burns, Brad Chapman, Tim Clark, Lee-Ann Coleman, Jay Copeland, Sudeshna Das, Antoine de Daruvar, Paula de Matos, Ian Dix, Scott Edmunds, Chris T Evelo, Mark J Forster, Pascale Gaudet, Jack Gilbert, Carole Goble, Julian L Griffin, Daniel Jacob, Jos Kleinjans, Lee Harland, Kenneth Haug, Henning Hermjakob, Shannan J Ho Sui, Alain Laederach, Shaoguang Liang, Stephen Marshall, Annette McGrath, Emily Merrill, Dorothy Reilly, Magali Roux, Caroline E Shamu, Catherine A Shang, Christoph Steinbeck, Anne Trefethen, Bryn Williams-Jones, Katherine Wolstencroft, Ioannis Xenarios, and Winston Hide. 2012. Toward interoperable bioscience data. *Nature Genetics* 44, 2 (Jan. 2012), 121–126. doi:10.1038/ng.1054
- [5] Hannes Ulrich, Ann-Kristin Kock-Schoppenhauer, Noemi Deppenwiese, Robert Gött, Jori Kern, Martin Lablans, Raphael W Majeed, Mark R Stöhr, Jürgen Stauber, Julian Varghese, Martin Dugas, and Josef Ingenerf. 2022. Understanding the Nature of Metadata: Systematic Review. *J Med Internet Res* 24, 1 (11 Jan 2022), e25440. doi:10.2196/25440
- [6] Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Olavo Bonino da Silva Santos, Philip E. Bourne, Jildau Bouwman, Anthony J. Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott C. Edmunds, Chris T. A. Evelo, Richard Finkers, Alejandra N. González-Beltrán, Alasdair J. G. Gray, Paul Groth, Carole A. Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A. C. 't Hoen, Rob W. W. Hooft, Tobias Kuhn, Ruben G. Kok, Joost N. Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel Laerte Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene C. van Schaik, Susanna-Assunta Sansone, Erik Anthony Schultes, Thierry Sengstag, Ted Slater, George O. Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik M. van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katy Wolstencroft, Jun Zhao, and Barend Mons. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3 (2016). https://api.semanticscholar.org/CorpusID:8755162
- [7] Katherine Wolstencroft, Stuart Owen, Olga Krebs, Quyen Nguyen, Natalie J Stanford, Martin Golebiewski, Andreas Weidemann, Meik Bittkowski, Lihua An, David Shockley, Jacky L. Snoep, Wolfgang Mueller, and Carole Goble. 2015. SEEK: a systems biology data and model management platform. *BMC Systems Biology* 9, 1 (July 2015). doi:10.1186/s12918-015-0174-y