Addressing the Research Metadata Bottleneck: Surveying Features for Engagement and Productivity

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

High-quality metadata is foundational to effective data discovery, management, and reuse. However, metadata documentation remains a bottleneck in research workflows, often perceived as tedious and undervalued [2]. This challenge is particularly acute in interdisciplinary domains, where metadata plays a critical role in ensuring data interpretability and usability. When researchers disengage from metadata tasks due to poor tooling or lack of motivation, documentation is often minimal or inconsistent, compromising data quality and creating challenges for downstream processes such as integration, analysis, and reuse. To investigate possible solutions to this problem, we developed a survey to evaluate how researchers perceive proposed features of metadata management tools, specifically in terms of their potential to enhance workflow productivity and engagement. These features target sociotechnical aspects, including motivation, recognition, and collaboration, as well as AI-powered guidance, metadata generation, and validation. Tools perceived as improving productivity and engagement are hypothesized to be more likely to be adopted and used consistently, which in turn leads to improved metadata quality and more reusable datasets. This survey is structured around the ISA framework (Investigation, Study, Assay) [3, 4], a widely adopted metadata framework in research that promotes interoperability and reusability across research domains. By grounding the survey in this framework, we aim to ensure that the features evaluated are relevant to real-world metadata practices and align with established standards. Additionally, we invite respondents to suggest additional features. By analyzing responses, we aim to identify which proposed tool features are most likely to reduce friction in metadata workflows, and collect additional features from researchers for assessment. Furthermore, the survey also collects information about professional background, including role, field of study, years of experience, and types of data typically handled. This allows us to explore whether certain features are perceived as more valuable by specific user groups. These insights can inform the development of more targeted and adaptive metadata tools that align with the workflows and expectations of diverse research communities. The survey will contribute to the development of the MetaBuddy, an AI virtual assistant, and the Resilience Hub, a research data management platform, by informing on highly rated features and new suggestions. These are being designed to support research projects conducted in CropXR, an institute based in the Netherlands dedicated to developing more resilient crops through data-driven plant research [1]. While rooted in this use case, this work contributes to the broader conversation on human-centred data management, offering insights into how user feedback can inform the design of intelligent metadata systems that promote FAIR principles [5].

References

[1] CropXR: Making crops more resilient, sustainable, and climate-adapted. `https://cropxr.org/`, 2025.

[2] Kaiser, K. A., Urberg, M., Johnsson, M., Kemp, J., Meadows, A., and Paglione, L. An international, multistakeholder survey about metadata awareness, knowledge, and use in scholarly communications. *Quantitative Science Studies 2*, 2 (07 2021), 454–473.

[3] Sansone, S.-A., Rocca-Serra, P., Field, D., Maguire, E., Taylor, C., Hofmann, O., Fang, H., Neumann, S., Tong, W., Amaral-Zettler, L., Begley, K., Booth, T., Bougueleret, L., Burns, G., Chapman, B., Clark, T., Coleman, L.-A., Copeland, J., Das, S., de Daruvar, A., de Matos, P., Dix, I., Edmunds, S., Evelo, C. T., Forster, M. J., Gaudet, P., Gilbert, J., Goble, C., Griffin, J. L., Jacob, D., Kleinjans, J., Harland, L., Haug, K., Hermjakob, H., Sui, S. J. H., Laederach, A., Liang, S., Marshall, S., McGrath, A., Merrill, E., Reilly, D., Roux, M., Shamu, C. E., Shang, C. A., Steinbeck, C., Trefethen, A., Williams-Jones, B., Wolstencroft, K., Xenarios, I., and Hide, W. Toward interoperable bioscience data. *Nature Genetics 44*, 2 (Jan. 2012), 121–126.

[4] Sansone, S.-A., Rocca-Serra, P., Gonzalez-Beltran, A., Johnson, D., and ISA Community. Isa model and serialization specifications 1.0.

[5] Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. The fair guiding principles for scientific data management and stewardship. *Scientific Data 3*, 1 (Mar. 2016).