Open Scholarly Metadata for Sharing the Nuances of Research Processes

Daniel Mietchen[1]

[1] *Leibniz Institute for Information Infrastructure*

Purpose

Various shades of openness have significantly improved the accessibility of research outputs. However, the research process itself, including granular details of study design, data collection, analysis, and interpretation, remains comparatively opaque in terms of the metadata available about it and options to find relevant research based on such metadata. This opacity is a barrier to collaboration and reproducibility, limits quality assessment, and impedes knowledge synthesis as well as educational efforts. This contribution argues that open access to research process information is crucial for advancing open science and proposes a framework for improving metadata practices in these under-documented areas. This directly addresses the workshop theme of "Open Access of Research Information" by focusing on a currently under-served area: the research process itself.

Method

Our approach combines a conceptual analysis of research workflows with an examination of current metadata practices and emerging standards. We recognize that research processes vary significantly across disciplines, but common decision points exist where enhanced documentation can significantly improve transparency and reproducibility. Our methodology involves the following steps. First, we deconstructed typical research workflows across diverse fields, identifying key decision points where researchers make choices that shape their findings and, in turn, the latter's reproducibility. These decision points, often implicit, represent crucial junctures where capturing and sharing metadata can be beneficial. For instance, ethical review boards require researchers to navigate complex ethical considerations and obtain necessary approvals. While the specific details can frequently not be shared for various reasons, it would often be possible to share categorical information regarding what aspect of the research triggered the need for ethical review, what kind of permissions were sought by the researchers, what kind of changes were requested during the ethical review, and what was the timeline of the ethical review process. Other decision points concern matters like the precise research questions to be addressed, the operationalization of measures for the variables occurring in these research questions, and choices regarding statistical approaches for defining study parameters like sample size or for analyzing the gathered data. Metadata should document these choices, explaining the rationale behind them, outlining any assumptions made, and detailing how these assumptions were checked. If a t-test is used, for example, metadata should specify whether it was a one-tailed or two-tailed test and record the specific version of the statistical software employed. For more complex analyses, the full code used should be shared, ideally in a public repository. Qualitative analysis, on the other hand, while often perceived as less structured, also involves crucial choices. Researchers make decisions about coding schemes, thematic analysis methods, and the interpretive frameworks they employ. Metadata should capture these choices, including detailed

codebooks, documentation of coding decisions, and the rationale behind interpretations. Documenting the inter-coder reliability for a qualitative coding scheme, for example, is essential for demonstrating the rigor of the analysis. Second, we reviewed existing metadata schemas and standards across different disciplines to identify best practices and areas for improvement. This review included examining domain-specific metadata standards (e.g., Dublin Core, DDI) as well as general-purpose metadata frameworks. While such standards exist in many fields for capturing many aspects of research processes, the corresponding metadata about individual research processes are often not shared in a way that would allow to find past or ongoing research based on just such metadata - for instance, if one team is working on a study for which it is difficult to reach a useful statistical power, they might want to collaborate with others working with similar data. Third, based on the analysis of research workflows and current metadata practices, we propose a framework for capturing and sharing metadata about research processes. This framework focuses on capturing the choices researchers make at each decision point, along with the rationale behind those choices. It builds on established practices for sharing research by experimenting with new publication types such as formalized definitions, hypotheses or research questions, and it emphasizes the use of standardized vocabularies and ontologies where possible to ensure interoperability. The framework is designed to be adaptable to different research disciplines and methodologies, recognizing the diversity of research practices. It aims to provide researchers with a practical tool for documenting their research processes in a clear, consistent, and accessible manner.

Results

Our preliminary findings indicate a significant gap in practices related to the sharing of metadata about research processes. Existing metadata schemas primarily focus on outputs - primarily publications, but also increasingly datasets or software - neglecting the granular details of how research is conducted. Our exploration revealed that researchers often rely on informal, non-standardized methods for documenting research processes. These findings highlight the need for a standardized framework for sharing metadata about research processes.

Value

This contribution offers a concrete framework for improving metadata practices related to the research process, with far-reaching benefits for open science and research integrity. By making the "how" of research more transparent, this framework promises to fundamentally enhance reproducibility. Detailed documentation of research processes allows for independent verification of findings, strengthening the foundation upon which scientific knowledge is built. Furthermore, increased transparency in research processes fosters greater trust in research outcomes. When the steps taken to arrive at conclusions are clearly documented and accessible, the public and the scientific community can have greater confidence in the validity of research findings. Beyond individual studies, this framework facilitates knowledge synthesis. Standardized metadata about research processes enables researchers to compare and integrate findings across different studies, accelerating the pace of discovery. Moreover, documenting negative results and failed experiments, a practice often neglected, can significantly reduce research waste by preventing duplication of effort. When researchers can easily access information about what has already been tried, even if

unsuccessful, they can avoid repeating unproductive lines of inquiry. Ultimately, this work contributes directly to the workshop theme by focusing on a crucial yet under-explored aspect of open access: the research process. The proposed framework offers a practical approach to improving metadata practices and promoting greater transparency and rigor in research, with broad implications for the future of open science.