FEATURE

Scholarly Metadata as Trust Signals: Opportunities for Journal Editors

Madhura S Amdekar

In recent years, research integrity issues are in the limelight with the emergence of new and complex threats, such as paper mills, citation cartels, fabricated peer reviews, fake papers, artificial intelligence–generated images, among many others.¹⁻⁴ A worrying feature of these emerging research integrity threats is that they often occur at scale and can affect many journals and articles at the same time. Taken together, this poses a considerable challenge to journal editors and editorial offices, which are key stakeholders in ensuring the integrity of the work they publish.

Scholarly metadata is an important tool that can be used in the endeavor to protect research integrity, especially to uphold the integrity of the scholarly record. The term "scholarly record" refers to the complex and interconnected network of published outputs (e.g., journal articles, books), the inputs that go into the creation of these outputs (e.g., datasets, preprints), and the metadata for these outputs.⁵ Preserving the integrity of the scholarly record is important because the scholarly record provides the foundation on which the global scholarly community can continue to build. When relationships between research outputs are not explicit, or when the metadata about these outputs are either incomplete or outdated, there is a risk that the scholarly community will not be able to access the most up to date information.

Metadata provide critical context about published works.⁵ By doing this, it acts as a marker for trustworthiness. Crossref provides infrastructure that allows those who create scholarly outputs to provide metadata for these outputs. These metadata are openly available; hence, others can use and evaluate them as evidence of the credibility of the

Madhura S Amdekar (https://orcid.org/0000-0003-1741-6646), is with Crossref.

Opinions expressed are those of the authors and do not necessarily reflect the opinions or policies of the Council of Science Editors or the Editorial Board of Science Editor.

https://doi.org/10.36591/SE-4704-10

published work. We at Crossref have been engaging with our community to find out which key metadata elements are the most pertinent as trust markers. This will help in informing the kind of scholarly communications infrastructure the community needs. We trust it will also encourage editors and journals to champion for the collection and deposition of these metadata elements.

Metadata as Trust Signal

It should be noted that the presence of a persistent identifier for a scholarly output in itself is not a signal of quality—a digital object identifier (DOI) only indicates that the item exists. Rather, it is the presence (or absence) of metadata associated with the scholarly output that serves as evidence of how the creators, or stewards, of these items ensure the quality of their content.

Crossref is an organization that enables its members to create and register persistent identifiers (Crossref DOIs) for their publications and to cross-reference the works. When members deposit their content with Crossref, they provide the basic bibliometric metadata related to scholarly publications, such as article title, publication date, names of authors, journal title, name of the conference, and volume or issue number.6 We also recommend the inclusion of authors' affiliations (preferably as Research Organisations Registry [ROR] IDs-more about those later) and ORCiDs, the abstract, and the list of references-all of which assist discoverability, as well as provide better context about the work at hand. Additional information, such as funding metadata, the relationships between objects (e.g., "is preprint of", "is review of", "is funded by"),6 and clinical trial numbers (where relevant), offer even more insight into how that scholarly output came about. The presence of this information means the metadata can tell us who authored a particular work, who funded it, where was the research carried out, what the relationship is between this work and a particular dataset, and so much more. Works such as journal articles may undergo changes even after publication. Additional information may be added to them, they may undergo a correction, or they might be retracted. Metadata that signifies these updates are essential in maintaining the

۲

CONTINUED



۲

Figure 1. Case example of an article record with rich metadata (Credit: E Atoni).

integrity of the scholarly record as they ensure scholars are reading and citing the most up-to-date work.

All the metadata that Crossref members provide in this way are made openly available through our application programming interfaces, in a machine-readable format, which allows downstream services and users to access this information to build tools and assess the trustworthiness of any research output. The more metadata elements associated with a research output, the more complete the information available to the community is for determining its trustworthiness (Figure 1). On the contrary, when metadata are missing, it is harder for others in the community to understand the context and therefore the credibility of the work.

When metadata are associated with a persistent identifier, a large database may be created that increases the feasibility of analyzing data at scale. For instance, ROR IDs, which are persistent identifiers for research organizations, can help in connecting problematic manuscripts to the institution where they were produced.⁷ An example of an advantage of persistent identifiers is the ability to identify organizations associated with paper mill articles.⁷

Working With the Community

Ever since its inception, all of Crossref's initiatives have focused on preserving the integrity of the scholarly record. In the past couple of years, we have made dedicated efforts to engage with the scholarly community on this issue. Our purpose has been manifold: as a community-led organization, not only did we want to share our thinking about the role we can play in the community to bring everyone together, we also wanted to get the community's input on our role and understand its perspective on leveraging metadata for trustworthiness.

()

 (\bullet)

۲

FEATURE

CONTINUED



Figure 2. An example of a grant linked from a journal article to the funder (Credit: G. Hendricks).

So far, we have held discussions with our community in 2022 and 2023, to which we invited publishers, research integrity experts, researchers, policy-makers, funders, editors, and organizations such as the Committee on Publication Ethics and STM.^{8,9} The focus of the rest of this article will be on 2 of the key questions that were asked of the participants: 1) What metadata are important for signaling trust? 2) What metadata are "nice to have" in the scholarly record?

Metadata Important for Signaling Trust

When it comes to signaling trust, information about retractions and corrections, references, abstracts, affiliations, and funding make it to the top of the list for our members. This is expected, given there have been several instances in the recent past where hundreds of manuscripts had to be retracted due to identification of concerns. Retractions, corrections, and other kinds of postpublication updates are crucial in keeping the scholarly record up to date and are evidence for readers that they are reading or citing the most recent version. References and abstracts are especially useful, as they can be subjected to text mining to uncover any systematic inconsistencies-again, an avenue where large scale datasets help to uncover large scale trends and tie them back to the concerned authors, institutions, and funders. With affiliations, authors can be consistently associated with their organizations, which is immensely helpful for circumstances in which ethical investigations need to be carried out, and organizations need to be involved. Grant identifiers and funder identifiers, which are part of funding metadata, make the link between research and its funders explicit (Figure 2). This enhances the trustworthiness of the research and helps to identify any conflicts of interest.

Metadata Horizon Scanning

The community also has several suggestions for metadata elements that are "nice to have" in the scholarly record for ascertaining trustworthiness. Given there have been increasing concerns about the quality of work reported in guest-edited issues, the need for more information around these special issues is a recurring theme. Identity of the guest editors, identity of reviewers, and submission and acceptance dates of the articles are the key pieces of information about special issues that the community would like to see. In general, more information about peer review, such as key dates in the process (e.g., submission, revision, acceptance) and the identities of the people involved (e.g., handling editor, reviewers, corresponding author), even for articles in regular issues, are welcomed.⁹ In addition to calls for registering retractions, community members have also called for addition of reasons for retractions and submission dates of retracted papers in the associated metadata. Metadata that add transparency to the research process, such as details of ethical approvals, clinical trials, and conflicts of interest are also deemed valuable by the community in this context.⁹ However, not all of these pieces are possible to be included in the metadata, and including expanded metadata related to editors is not a part of our roadmap currently. We are working on several metadata development projects, many of which are tied to the integrity of the scholarly record (read more about our plans¹⁰).

Opportunities for Journal Editors

Upholding the integrity of the scholarly record requires combined efforts from everyone in the scholarly community. Editors and editorial offices are at the forefront of the endeavor to uphold the integrity of the work they publish. Editors drive the day-to-day editorial

۲

()

30/01/2025 02:42:01

 (\bullet)

۲

CONTINUED

and production operations of journals, including the submission of their publications' metadata to Crossref. As outlined above, metadata are not just a part of the production workflow but also are important in signaling the integrity of scholarly output. Editors have a crucial role to play in this effort. By virtue of their close relationship with authors, peer reviewers, and editorial boards, editors are uniquely positioned to gauge the metadata that are relevant for signaling trustworthiness in a specific subject discipline. They can champion expansion of facilities for capturing such key metadata in their submission process with colleagues and vendors. Editors can encourage prospective authors and researchers to deposit key metadata with their submissions and to adopt journal policies that allow publishers to collect and deposit this metadata with Crossref. Crossref also provides a community forum for editors to share their knowledge and best practices with one another in this domain.¹¹ There is much work to be done on this front. At Crossref, we are very keen to hear more from the community about how we can work together to preserve the integrity of the scholarly record. We welcome your thoughts on the key metadata elements that you think signal trust and how it can be made easier to collect and deposit this metadata.

Acknowledgements

I thank Kornelia Korzec for her review and feedback, which significantly improved the article. I thank Evans Atoni and Ginny Hendricks for providing Figures 1 and 2, respectively.

FEATURE

References and Links

- https://publicationethics.org/resources/forum-discussions/ publishing-manipulation-paper-mills
- Fister I Jr, Fister I, Perc M. Toward the discovery of citation cartels in citation networks. Front Physics 2016;4:49. https://doi. org/10.3389/fphy.2016.00049.
- https://www.chemistryworld.com/news/review-mills-identifiedas-a-new-form-of-peer-review-fraud/4018888.article
- Gu J, Wang X, Li C, Zhao J, Fu W, Liang G, Qiu J. AI-enabled image fraud in scientific publications. Patterns (N Y). 2022;3:100511. https://doi.org/10.1016/j.patter.2022.100511.
- 5. https://www.crossref.org/blog/isr-part-one-what-is-our-role-inpreserving-the-integrity-of-the-scholarly-record/
- Hendricks G, Tkaczyk D, Lin J, Feeney P. Crossref: the sustainable source of community-owned scholarly metadata. Quant Sci Studies 2020;1:414–427. https://doi.org/10.1162/qss_a_00022.
- 7. https://ror.org/blog/2023-11-27-clear-skies-case-study/
- 8. https://www.crossref.org/blog/isr-part-four-working-together-asa-community-to-preserve-the-integrity-of-the-scholarly-record/
- 9. https://www.crossref.org/blog/frankfurt-isr-roundtableevent-2023/
- https://www.crossref.org/blog/metadata-schema-development-plans/
 https://community.crossref.org/

A