

PO-04

Open infrastructure and scholarly metadata for research integrity

Madhura Amdekar

Crossref, Netherlands

With an increasing volume of research being published every year, the scholarly record continues to grow both in terms of the number of outputs and their diversity. Preserving its rectitude is of paramount importance given the research integrity problems that continue to affect research publications at scale. Scholarly metadata has an important role to play in this endeavour. Complete and accurate metadata acts as “trust signals” by answering questions about scholarly works such as who funded the work, where it was published, was it corrected or retracted, and thereby provides important context against which the integrity of the work can be assessed. Therefore, metadata has a much wider role to play than the commonly perceived notion of it being only a part of publishers’ and journals’ production workflows for increased discoverability. Rather, when the right metadata trust signals are available, the scholarly community can use them to identify scholarly trends, including for identifying problematic publishing practices.

Open infrastructure providers, such as Crossref, enable its members to register metadata about the works that they publish. Crossref is a not-for-profit membership organization with nearly 22,000 members from 160 countries, and most members identifying themselves as research institutions. Crossref assigns DOIs and collects metadata for 30 research output types, some of which are journal articles, preprints, and datasets. The collected metadata ranges from elements such as author names, titles, publication dates, to metadata on funding, retractions and corrections, relationships between works (“e.g. is preprint of”) and more. All this metadata is made openly available via Crossref’s API and used by several downstream services such as OpenAlex, Dimensions, Scopus, Web of Science, and others that are used by researchers and across the wider scholarly community.

In this presentation, we’d like to introduce the audience to scholarly metadata and their role as trust signals in assessing the trustworthiness of research outputs. We will provide an overview of the open scholarly infrastructure, talk about the metadata that is made available by Crossref, and focus on elements such as author affiliations and ROR IDs, funding metadata, retractions and corrections, which should be of interest to this audience. We will discuss how open metadata can help institutions to track the impact and integrity of their outputs, including tracking where the research was published, what research was retracted, what is the source of funding for publications, etc. We will share examples of how others in the community have used the openly available metadata to identify problematic scholarly content. We hope that this will inspire the attendees to incorporate collection and deposition of metadata as a component of good research integrity practices at their institutions, and leverage the presence (or absence) of metadata as another evidence of integrity.