

Metadata and discoverability: A use case overview

Jennifer Kemp*

Head of Business Development, Crossref, 50 Salem Street, Lynnfield, MA 01940, USA

Abstract. This paper presents a brief overview of open interface options and use cases for Crossref metadata. The collective effect of the community's use of metadata on content discoverability—and the need for richer metadata to support research communications—are presented in the context of a greater focus on non-bibliographic metadata. An increase in metadata querying and the core operating principle of transparency are discussed as aligned factors in making open, high-quality metadata a strategic priority for Crossref.

Keywords: APIs, Crossref, discoverability, editors, infrastructure, libraries, Metadata, persistent identifiers, publishing, research, scholarly communications

1. Introduction

The use/reuse of metadata has become a topic of significant interest within the scholarly communication community in recent years, prompting much discussion and numerous presentations at industry conferences, for example. Crossref, which works with over ten thousand publisher members and a few thousand more metadata users of various services, now has nearly one hundred million metadata records. At this scale, and because the underlying infrastructure is largely hidden, it is mutually beneficial for Crossref and the research community to understand and engage in shared efforts to strengthen the scholarly record by making research outputs linked and discoverable with high-quality metadata that is open to allow reuse.

2. A renewed focus on metadata

While bibliographic or citation metadata has long been considered sufficient for describing a work - the core function of metadata - it might now be thought of as minimal. Among the elements most frequently requested by users of Crossref metadata are abstracts, authors' references, license information, funding information, and author affiliations. The 'name, rank and serial number' approach of bibliographic metadata often does not provide enough detail for systems that want the fullest picture possible.

Users seek a few key qualities in metadata beyond simply the inclusion of certain elements: timeliness, accuracy and completeness. In other words, how quickly metadata is made available may be considered

*E-mail: jkemp@crossref.org.

just as important as how much correct information is provided. Even basic elements such as publication dates receive considerable scrutiny by users' systems, which reasonably expect, but don't often receive, consistently formatted metadata. When considered in the context of discoverability, richer metadata makes sense. While a given metadata record stands on its own as a description of a work, in practice it is often presented alongside other metadata records. The more complete the records are, the more an individual record can stand out or be used in analyses with others.

3. Why open?

Discussing and communicating the value of metadata can be a challenge. It can be a technical, dry topic and, if metadata creators do not understand how metadata is used, it can be difficult to make the case for enriching it. Yet the whole scholarly community is involved in or affected by metadata, directly or indirectly. For these reasons, Crossref is described as a 'community' infrastructure and it is within this context that open channels for metadata are provided, to encourage use by a broad section of the community. Simple interfaces as well as very sophisticated ones are available, for all levels of technical ability and frequency of use.

4. Community use cases

Users of Crossref metadata come from all around the world and all types of organizations. The use cases and workflows vary widely. Publishers, for example, are sometimes themselves users of metadata beyond that which they themselves generate.

4.1. Simple examples

Simple but significant use cases often, but not always, involve human interfaces. Libraries, for example, often use OpenURL, bibliographic metadata that powers links resolvers, which in turn direct readers to content. The Crossref metadata web search provides each result in several reference formats, making it easy for an author, for example, to copy and paste a citation in APA, MLA, or other style formats. A common use case for metadata is citation-matching, or finding a DOI for a given reference. Editors often do this important work, as do machines, but in different ways. The former may use an interface known as Simple Text Query. Occasional, low volume, and new users may find it easiest to take a simple approach. Human interfaces may also be useful for publishers that need to see what is in their metadata quickly and without needing any technical expertise. But it can also be helpful to see metadata *in situ* and a few REST API records in JSON for this context are probably easy to tolerate, particularly with a JSON viewer browser extension.

4.2. Machine use of metadata

Those who work with large portions of the corpus and do a lot of querying use machine interfaces. The Crossref REST API is used in systems throughout scholarly communication for functions as varied as enhancing search services, applications of artificial intelligence, and reporting on funding and author activities. Reference managers, metrics providers, library vendors, and research organizations are among

the users and services. These organizations often report metadata quality issues because their systems and workflows may be significantly and negatively affected by inconsistencies in metadata, let alone poor quality.

In one 2017 conference session, to answer the question ‘What is the single most important thing publishers should know about how you use metadata?’ a librarian replied simply, “We reuse it.” The answer may be vague, but the fact that exactly how metadata is used cannot always be known upfront may serve as a good reminder that its message may be carried far and wide.

5. Metadata and discoverability

Crossref sees well over six hundred million metadata queries per month across all interfaces, up more than a quarter over the previous year, which itself was an increase from the year before that. It is difficult to think of so many queries and consider the breadth of systems and services that perform them and not conclude that metadata affects discoverability. In that light, it stands to reason that open interfaces and encouraging the use of its metadata can benefit the community.

When considered within the context of the scholarly record, it becomes clear that metadata elements must develop over time, with new elements introduced to accurately reflect the content being produced. Witness the increase in the number and types of persistent identifiers (PIDs), which now have their own conference, PIDapalooza. Identifiers, of course, are but one kind of metadata element. A few more recent changes highlight the need for ongoing attention to developing metadata as part of the scholarly record.

5.1. Examples

In 2016, Crossref introduced linked clinical trials - a small, but powerful example of a gap in the record being filled in a way that allows for systems to recognize research outputs like articles that resulted from clinical trials. That otherwise obvious connection was made easily available only two years ago.

Later that same year came preprints, which are linked in the metadata to versions of record, where they exist. This year, Crossref introduced peer review reports. Currently in the works are organization and grant identifiers. Event Data uses metadata to report where publications have been linked on the web – for example, from blogs, Reddit and Twitter.

These particular examples serve to illustrate that there is likely more to metadata than how it has been traditionally considered and that its role in the scholarly record is evolving.

6. Crossref resources

How to get Crossref metadata: <https://www.crossref.org/services/metadata-delivery/>

REST API: <https://api.crossref.org>

REST API documentation: <https://github.com/CrossRef/rest-api-doc>

Metadata web search: <https://search.crossref.org/>

OpenURL: <https://support.crossref.org/hc/en-us/articles/214880143-OpenURL>

Simple Text Query: <https://apps.crossref.org/simpleTextQuery>

Event Data: <https://www.crossref.org/services/event-data/>

About the Author

Jennifer Kemp is Head of Business Development at Crossref, where she works with organizations around the world that use metadata in systems and services throughout scholarly communications. Previously, she was Senior Manager of Policy and External Relations for Springer Nature. She is active in the Metadata 2020 initiative and co-chairs its Shared Best Practice and Principles project. E-mail: jkemp@crossref.org.