



Citation Needed: Metadata reach and return RFP

[Citation Needed: Metadata reach and return RFP](#)

[Project summary](#)

[Background on Crossref](#)

[Problem statement](#)

[Project goals](#)

[Approach](#)

[Timeline \(estimate\)](#)

[Sought qualifications](#)

[Our team](#)

[Select background resources](#)

[How to respond](#)

Project summary

At [Crossref](#), we often talk about the benefits of more and better metadata. This project is designed to provide an evidence-based answer to those who reply, “Prove it.”

There is a mismatch between discussions of the importance and benefits of quality metadata, evidence to support or clarify these assumptions, and resources needed to capitalize on the increasing attention to these and related questions. This recent, heightened focus on metadata makes now an opportune time to address these issues. Accordingly, we seek a consultancy or research group to analyze and report the results of broad-based research into any effect(s) of the availability, completeness, and quality of metadata in scholarly communications.

Background on Crossref

Crossref is a small, not-for-profit membership organization supporting scholarly communications and publishing by providing infrastructure for over [126 million](#) and growing metadata records for journals, books, grants, and more. We are the largest of a handful of DOI (Digital Object Identifier) [registration agencies](#) ensuring the persistence of linking to and among scientific research outputs. We operate with a [small staff](#) of 40 spread across the US and Europe and a [governing board](#) of 16 member organizations.

Crossref’s community has mirrored the changes in scholarly publishing and communications since our founding 20+ years ago. Our diverse, global, and rapidly growing membership of over 15,000 varies greatly in organization type, size, resources, business models, and geography



(over 140 countries). Crossref services have also grown and diversified, well beyond DOIs. Non-members are also a significant part of our community; they are equally diverse, including researchers, librarians, startups, funders, and a host of organizations providing tools and services powered by Crossref metadata and associated services.

Crossref metadata is very widely used by the broader global scholarly community, and improving it is a crucial part of our [strategic roadmap](#).

Problem statement

With the notable exception of the effects of [book metadata](#) on sales, there are surprising gaps in evidence of direct or indirect effects that accurate, enhanced metadata may bring to publishers or the wider world of research.

Scholarly metadata from a variety of sources, not just Crossref, is widely used in a multitude of systems and services in scholarly communications and is frequently discussed in library contexts. However, there is a need for analyses and published research that provides specific guidance or evidence for evaluating the effects of metadata and determining needed improvements.

In addition, discoverability, quality, and completeness are all terms open to interpretation in the context of metadata and may be affected by issues of distribution and access.

Project goals

This project is intended to provide:

1. Evidence of metadata and its effects
2. A set of resources for the community to build upon

The outputs should include, but need not be limited to:

- Identifying one or more methodologies for researching the topic(s)
- Defining the key terms as used in the research context
- A framework for making practical changes and assessing their results
- Publishing the data collected
- Suggesting any future work needed
- Publishing a final report of the work, as a paper, preprint, report, etc. (tbd)

Approach

Given how broad, varied, and numerous the systems and the metadata are, the suggested approach is one of sampling. Original and data-focused research seems warranted, possibly in combination with other approaches, such as a literature review, interviews, etc.

The following list is not exhaustive but provides candidates to consider sampling from:

- metadata elements
- subject disciplines
- content types (books, journals, grey literature, etc.)
- systems and tool types (library discovery systems, recommendation services, etc.)

Additional factors to consider are:

- Open Access vs. subscription/purchased content
- Free vs. fee-based metadata sources
- Errors in metadata
- Size of the corpus (metadata or content)
- Frequency and types of metadata updates over time
- Accessibility of metadata

Analyses should address:

- Differences in bibliographic and non-bibliographic metadata
- Content with related outputs, for example a publication and its associated data, a publication and any translations, etc. that are linked in metadata vs. those that are not, for example versions of record not linked back to preprints
- Measuring 'before and after' changes of enhanced or corrected metadata

Specific samples, questions, instruments, and audiences to be discussed.

Timeline (estimate)

The scope and methodology of the research should guide the overall timeline but the project is work expected to take about a year, with the total timeline accounting for holidays:

- July 2021: RFP announced
- Early September: RFP responses due
- Early October: Partner group selected
- Mid-December: Project scoped and reviewed
- January 2022: Project plan approved, research begins



- Monthly check-ins with Crossref through mid-2022
- Summer 2022 check-in: Research complete, analysis begun, review plan to finalize work
- 4Q22: Write results, publish (or submit for publication) and publicize results

Sought qualifications

Knowledge of or familiarity with scholarly communications, publishing, and/or metadata is not necessary but helpful. We seek an agency, consultant, research group, or analyst with experience in the following areas:

- Qualitative and quantitative research methodologies
- Querying REST APIs and working with JSON and other data formats
- Ability to synthesize varied and large datasets
- Communicating research to professional and non-specialist audiences

Our team

[Jennifer Kemp](#) is the lead on the project and will be the day-to-day contact. [Ginny Hendricks](#) is the sponsor for the project and together with Jennifer will select and partner with those doing the research. Other team members may be brought in as needed such as Member Experience, Community Outreach, or other members of the management team. We are a committed, fun and flexible group and enjoy working with smart, creative people who use a collaborative, data-driven approach to community research.

Select background resources

[Metadata 2020](#) literature review: <https://doi.org/10.3897/rio.5.e38698>

A few general sources

- [Metadata Effectiveness in Internet Discovery: An Analysis of Digital Collection Metadata Elements and Internet Search Engine Keywords | Yang](#)
- [A novel framework for assessing metadata quality in epidemiological and public health research settings](#)
- [Metadata & discoverability](#)
- [Collaborative Improvements in the Discoverability of Scholarly Content](#)
- [Discoverability in \(a\) Crisis](#)

A few sources specific to books

- [Shared Values, Common Benefits: The ScienceOpen OA Book Metadata Project funded by the German Federal Ministry of Education and Research](#)



- [Nielsen Book UK study: the importance of metadata for discoverability and sales](#)
- [Nielsen Book US study: the importance of metadata for discoverability and sales](#)
- [The Importance of Metadata for Discoverability and Sales](#)
- [Moving Metadata From How to Why](#)

How to respond

Responses should be submitted by 1st September to [Jennifer Kemp](#) and should include:

- A statement of interest in the project
- An outline of the proposed approach to the work
- Separate, itemized cost estimates, including for any proposed phased work
- Deliverables, including file formats
- Methodologies for all research and data collection and analysis
- Evidence of data integrity and security practices, e.g. GDPR
- A primary contact as well as a description of the project team
- Examples of similar work, if available
- References from previous work
- Any questions, deliverables or contingencies necessary from Crossref to meet milestones and budget

Thanks so much for reading. We look forward to hearing from you!