The research nexus and Principles of Open Scholarly Infrastructure (POSI): sharing our goal of an open, connected ecosystem of research objects

Rachael Lammey
Crossref, Oxford, UK

Abstract
As a community, it is impossible to ignore the fact that sharing research and information related to research is a much broader proposition than sharing an article, book, or conference paper. In supporting an evolving scholarly record, making connections between research organizations, contributors, actions, and objects helps give a more complete picture of the scholarly record, which open infrastructure organizations like Crossref call the research nexus. Crossref is working to support this evolution and is thinking about the metadata it collects via its members and that it supplements and curates, to make it broader than the rigid structures traditionally provided by content types. Furthermore, because of Crossref’s commitment to the Principles of Open Scholarly Infrastructure (POSI), this network of information will be global and openly available for anyone in the community to access and reuse. The present article describes this vision in more detail, including why it is increasingly important to support the links between research and elements that contribute or are related to that research; how Crossref, its members, and the wider community can support it; and the work and planning Crossref is doing to make it easier to achieve this.

Keywords
Research nexus; Metadata; Principles of Open Scholarly Infrastructure (POSI); Crossref; Identifiers

Introduction

Background
As of July 2023, Crossref holds a metadata store of over 146 million records, containing information and identifiers for journal articles, grants, peer reviews, preprints, and more. This metadata is widely and intensively used by the research community, with over one billion monthly
queries hitting Crossref’s metadata retrieval services, meaning that the information on the records registered by Crossref members disseminates throughout the research ecosystem. This dissemination is also increasingly important as Crossref’s membership is diversifying, extending to more countries and across more types of organizations. For example, universities and research institutions are now Crossref’s largest member type, joining library publishers, scholar publishers, and funders who are also growing in number.

In parallel, open infrastructure and identifiers are being created to help establish more granular persistent links in the research ecosystem. Research Organization Registry (ROR; https://ror.org/) IDs for organizations and grant identifiers are important metadata elements that support funding and publication workflows, in turn showing the reach and return on grants and awards and adherence to funder policies, especially those focused on open science, including open access.

A critical component of providing the infrastructure to support this information, however, is that the community needs to be able to trust that infrastructure and know that the information will be openly available for the long term. This is why Crossref is publicly accountable to the Principles of Open Scholarly Infrastructure (POSI) [1] practices of governance, insurance, and sustainability and is making the commitment to better meeting these practices over time, alongside other like-minded organizations such as Directory of Open Access Journals (DOAJ), Europe PubMed Central (PMC), and Connecting Repositories (CORE) [2].

**What Can a Network of Metadata and Relationships Help Achieve?**

The research nexus is a vision to which Crossref aspires: a rich and reusable open network of relationships connecting research organizations, people, things, and actions. It provides support for a number of areas that the research community sees as critically important.

Research integrity has long been a focus of what Crossref does, most clearly demonstrated by the Similarity Check service that provides the iThenticate tool alongside a comprehensive database of scholarly content so that Crossref members can check submissions for originality. Beyond this, the metadata registered by Crossref members helps provide signals about the trustworthiness of the work including provenance information, such as who funded it, which people and organizations contributed, whether something was updated or corrected, perhaps with a retraction notice, expression of concern, or additional supplementary materials.

Discoverability has always been at the core of why organizations register content with Crossref. It makes it easier for anyone interested in the research to find it consistently over time, and having a store of rich metadata all in one place makes it easier for the thousands of systems that use Crossref metadata to ingest it and combine it with other data, tools, and services. A researcher might come to search the literature via many different entry points—the Open Researcher and Contributor ID (ORCID) or name of an author or investigator, the funder, the organization, the publication name, or the license a piece of work can be reused under. The research nexus supports making and discovering work via these routes, provided that the relevant metadata has been registered by the member or supplemented by Crossref.

Reproducibility is closely related to research integrity. Providing or adding relationships in the Crossref metadata to link literature, data, software, protocols, and more, can provide transparency and give context to the findings and processes that underpin any research output.

Finally, this network of information can help with reporting and assessment. Organizations such as universities, funders, and governments need to track and demonstrate the outcomes of their investment, show compliance with funder mandates, and inform their strategies such as deciding what other research to fund. This kind of information can and should be included in Crossref metadata to support this work (Fig. 1) [3].

**The Importance of POSI**

The areas described in the previous section are essential for the research community. The foundations of that information are also necessary, and need to be a shared resource that the community can rely on for the long term or until they are no longer needed. This is why in November 2020 the Crossref Board voted to adopt POSI. This means that Crossref has publicly placed importance on several key areas: governance, sustainability, and insurance. Broad community governance means that an organization is governed in a way that is representative of its membership, so it can be steered in a direction that serves its stakeholders and can respond to the needs of the community. This is especially important as those stakeholders can grow in breadth and diversity over time, as exemplified by funders joining Crossref specifically to register grants.

The organization itself also needs to be resilient. Crossref needs to be sustainable so that it can fulfill its mission and have a contingency fund underpinning it which can be used if needed. For Crossref, the “insurance” principle under POSI means openness. Making sure the metadata is open and the code for Crossref services is increasingly open and forkable means that the community is not “locked-in” if they think that Crossref is not serving their needs and they need to take their work in another direction. An added advantage is that the community
can also contribute to Crossref’s code, cocreating the infrastructure, tools, and services that can be used more broadly. It also means a commitment to transparent operations. Publishing policies, practices, and documentation means that the community can see the details of what Crossref is doing and why. It also means that Crossref publishes its progress towards fulfilling the POSI principles [4] so that the community can see the areas where it still wants to improve.

The POSI principles can also serve as a decision framework for evaluating new projects to work on (e.g., they need to be open source and based on open data), and also new partnerships, prioritizing those with other organizations who publicly commit to POSI.

**Projects that Help Achieve the Research Nexus**

Crossref has long advocated for its members to provide comprehensive metadata when they register their records, including supporting complementary industry projects like Metadata 2020 and the Initiative for Open Citations (I4OC). However, it is also important that Crossref does its own work to support its members in providing this information and the community in using it. As Crossref outlines in its strategic agenda and roadmap [5], there are a number of existing and upcoming projects that aim to build out the research nexus.

Crossref will continue adoption activities to focus on top metadata adoption priorities: the deposit of references, abstracts, grants, ROR IDs, and data citation. In 2022, Crossref started adding ROR IDs registered in the metadata to its representative state transfer (REST) application programming interface (API) [6]. As of July 2023, Crossref can see over 54,000 records, including grants, journal articles, peer reviews, and preprints that contain a ROR ID. This supports improving the accuracy and potential for reuse of affiliation information related to research, meaning that an institution could more easily come to Crossref to find the outputs produced by its researchers.

In the past, Crossref members were able to choose to keep their references closed, or only make them available to organizations who publicly pay for that portion of the metadata. The
The research nexus and POSI

POSI principles only support the generation of revenue via services, not via data. In support of this, Crossref opened all of the reference metadata registered by its members by default [7] so that they can be reused to maximum effect by the community to expose and explore relationships between research, for example how one piece of research has continued or contradicted the work of another, or patterns of citation related to works that have been retracted.

Crossref is also working to develop its metadata model so that it is increasingly clear who is asserting specific pieces of information, for example, where Crossref has made a digital object identifier (DOI) match to an associated record or the member has done so. These assertions could also come from external sources such as when the ROR registry asserts that a funder ID is equivalent to a ROR ID. This more flexible metadata model also aims to make it easier for anyone using the Crossref metadata to query it at a more granular level and via a larger range of entry points. The first test of this model is an early version of a relationships API endpoint [8] that combines event data with citation and relationship metadata from members. It reports whether a connection exists between two metadata records and is key to implementing the research nexus. This will be built out with a greater volume of Crossref metadata over time, and more functionality will be added so that users can use more filters to get the specific subset of information they are interested in, such as citations to related research data or software. Crossref will also undertake a larger project to revisit and expose its matching strategy and functionality, starting with improving how it identifies and provides information on suspected connections between preprints registered with Crossref and associated versions of record on publisher websites.

The metadata that Crossref collects is also key to making the connections. Crossref is working to develop a clear metadata development strategy and gather priorities from the community, as evidenced by a metadata survey conducted among Crossref members in early 2023. To support this, the Crossref Labs group is building a test environment for metadata schema updates so that Crossref members and service providers can test out what metadata updates look like and practice depositing that metadata before it is “live.” This work aims to help the early adoption of new metadata elements as they become available. Finally, Crossref’s main mechanism for metadata retrieval is the Crossref REST API, and improvements to it to ensure metadata can be delivered reliably and at scale are also ongoing.

There are, however, bigger societal challenges than metadata. Research is global in nature, and organizations should not be prevented from participating in the research nexus because of financial constraints. In December 2022, Crossref announced the launch of its Global Equitable Membership (GEM) program [9]. Based on the International Development Association (IDA) list [10] and excluding anywhere Crossref is bound by international sanctions, organizations based in countries listed in the GEM program will be eligible to join Crossref and contribute with their metadata to a robust scholarly record at no cost. This program is also applicable to 187 existing Crossref members (as of December 2022) in eligible countries who are no longer charged for their membership or content registration. Crossref aims to grow the adoption of the GEM program so that it can include more of the world’s metadata and support more members in providing it.

Conclusion

Research and scholarship continue to exist in a changing ecosystem. One element that is evolving quickly is that funders, institutions, researchers, and providers of tools are services playing a larger role in shaping the policies and practices that accompany the sharing, assessment, and publication of research and objects related to that research. Crossref must help its diverse membership by capturing more provenance, relationships, and identifiers to meet their needs, but also the needs of the wider community. Its systems and metadata schema and infrastructure need to support this as well. Underpinning this is the commitment to POSI, so that Crossref can continue to be well-guided and sustainable. With a more complete, open, and connected picture of the scholarly record available, everyone will be able to examine the integrity and outcomes of collective efforts to progress science and society.

Conflict of Interest

Rachael Lammey is the Director of Product at Crossref. No other potential conflict of interest relevant to this article was reported.

Funding

The author received no financial support for this article.

Acknowledgments

The author thanks Ginny Hendricks for her help in proofreading this paper.

Data Availability

Data sharing is not applicable to this article as no new data were created or analyzed in this study.
Supplementary Materials

The author did not provide any supplementary materials for this work.

References

3. Hendricks G. The research nexus [Internet]. Crossref; 2021 [cited 2023 Jul 18]. Available from: https://www.crossref.org/documentation/research-nexus/
5. Crossref. Strategic agenda and roadmap [Internet]. Crossref; [cited 2023 Jul 14]. Available from: https://www.crossref.org/strategy/
8. Rittman M. Relationships are here [Internet]. Crossref; 2023 [cited 2023 Jul 14]. Available from: https://community.crossref.org/t/relationships-are-here/3523