

The Research Nexus vision for a more connected scholarly community

Jennifer Kemp* and Patricia Feeney
Crossref, Lynnfield, MA, USA

Abstract. Crossref envisions “a rich and reusable open network of relationships connecting research organizations, people, things, and actions; a scholarly record that the global community can build on forever, for the benefit of society”. This Research Nexus expands on the importance of research objects being persistently and uniquely identified. The scholarly community has an established practice of connecting things such as citations to others’ work and it is increasingly critical to identify relationships beyond citations, bringing together published work, unpublished work, institutions, individuals, and identifying the actions that they take e.g., funding, publishing, creating, modifying, citing, and sharing. The Research Nexus brings together metadata and relationships to build a joined-up picture of the scholarly ecosystem and helps everyone identify these relationships and how they change through time. This vision is possible if all parts of the scholarly ecosystem (and beyond) work together, including various scholarly infrastructure organizations.

Keywords: Metadata, scholarly record, relationships, research nexus

1. Introduction

For the past several years, Crossref has been focused on getting members to fully describe their outputs in the metadata. Publishers overall are much more aware of the value of more and better metadata than in the early days of DOIs and the pandemic has highlighted the consequences of problem metadata [1]. Now is the time to capitalize on this attention to metadata and move towards a Research Nexus of outputs that are connected to other records and information. The idea is a significant expansion of well-established practices of identifying relationships of one output to another - citations being probably the most common. A few years ago, Crossref expanded the few relationships we had at the time (for example, declaring one item to be a translation of another) and developed a robust, flexible set of options to allow for outputs to be related not only to other Crossref records, but also to DataCite records as well, to support emerging data and software citation practices. There are now over one and a half million relationships among Crossref’s one hundred and thirty-five million records, but of course it is almost certain that there are many more relationships that could be established among such a large corpus.

This paper develops the Research Nexus vision, based on a talk given at the NISO Plus conference in February 2022.

2. What are metadata relationships

Scholarly infrastructure is constructed of connections between researchers, research data, funders, research outputs, and more. Relationships are a way to connect related digital objects and records with

*E-mail: jkemp@crossref.org.

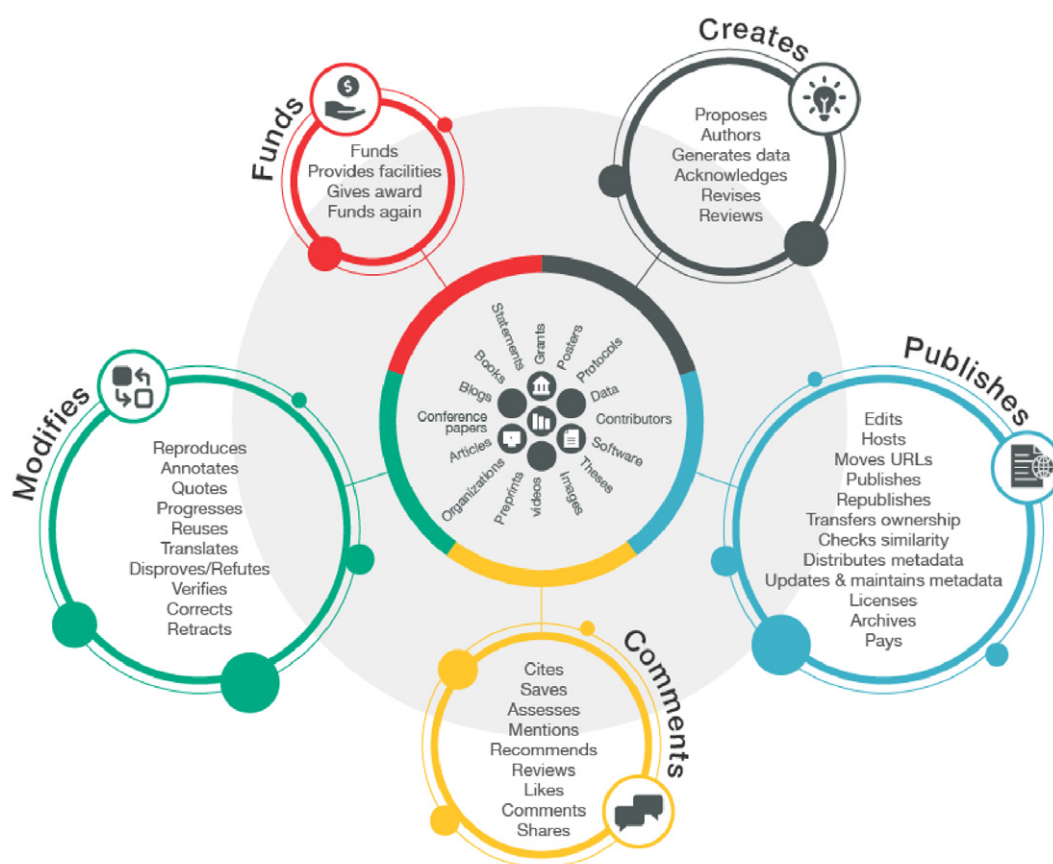


Fig. 1. Crossref's vision of the Research Nexus.

each other through metadata. Some relationships are well-established - references, for example, create a relationship between a citing article to other articles, datasets, software, and other resources used to support the research within an article - while others are increasingly made visible through explicitly-supplied relationship metadata. A preprint can be connected to an article, an article associated with a license, a translated work can be connected to the original, reviews can be connected to reviewed items, versions can connect to each other, and they all form the Research Nexus.

Relationship metadata is the foundation of the Research Nexus, and most metadata comprises relationships - a work is authored, edited, cited, reviewed, corrected, withdrawn, and more. Many of these relationships are shown in Fig. 1.

For example, a journal article can be connected with a dataset used in the research phase, or a preprint can be connected with a version of record (VOR). Members can also connect translations to each other, connect reviews to the works being reviewed, connect a research protocol to an article about the research, and of course connect a dataset to all of the posted and published research it generates.

Crossref enforces some relationships - preprints registered with Crossref must include relation metadata identifying the Version of Record (VoR) article when it is published. Our peer review content type also requires that the reviewed item DOI be included via relations.

We also connect items with corrections, updates, retractions, and clinical trials (using Crossmark), as well as grants and funding. Registering component records for figures, tables, and supplemental materials automatically relates those items to the parent article.

3. How relationships support the scholarly record

Reviews and preprints are currently among the most common relationships. Because of the requirement to connect preprints to VoRs where they exist, it is no surprise that these relationships are so common. VoR publishers may optionally link back to preprints using *has-preprint* and that is in the top five of relationships in the metadata. Registering peer reviews became an option just a few years ago and is among the fastest-growing content types registered with Crossref.

Translations have been available for a long time but are not used very often. It is hard to know if that is because of options such as including abstracts in multiple languages, for example, issues around multilingualism in the metadata or other reasons, such as a lack of awareness of the option.

Components, which include common items such as supplementary data and figures and tables, are among the most often registered type of record. They are also one where, like book chapters, information is inherited from the parent record. Even with the implicit *isChildOf* relationship, components should be explicitly connected to related records (and there can be more than one parent record) through use of *isPartOf*.

Data and software citations are a little bit different in that they can be expressed in reference lists as well as through *isSupplementedBy* relationships.

It is worth noting that even though many of the relationships mentioned here are fairly recent, these child-parent relationships, like citations, have been around for a long time. Including additional relationship types is a logical extension of long held practices and another example of the ways in which scholarly metadata continues to evolve.

Making relationships explicit and available in the open metadata makes it easier for users of the metadata, including machines, to establish connections among research outputs in an unambiguous and interoperable way, which goes a long way to realizing the Research Nexus vision. As with other metadata, the more there is in the metadata, the more consistently they are applied, the more the global user community of scholarly tools and services can rely on this information to support a wide variety of use cases for the scholarly record.

4. Moving toward the Research Nexus

It was not until fairly recently that there was enough non-bibliographic metadata to support relationships far beyond such things as traditional citations and parent-child connections. Event Data, introduced in 2016, remains a unique, open source of non-traditional, post-publication online commentary and is complementary to member-provided publication and grant records. Event Data, plus development of the deposit schema to expand relationships are cornerstones of the Research Nexus, which has led Crossref to consider changes to our API to make it easier to find and integrate relationships among records, including events.

5. Every community has a role to play

As with other metadata efforts [2], stakeholders across the community have a role to play in establishing relationships among records as a new norm. **Publishers and funders**, of course, are responsible for the metadata in their Crossref records, but since many use third party **service providers** in their supply chains, it is worth pointing out the necessity of having partner systems able to collect, deposit and/or distribute relationship information. Changes such as introducing new elements or refining workflows are a good opportunity to review what is and should be in the metadata, but existing and new records can have relationships added without requiring additional changes. Publishers considering adding relationships to metadata might start with a review of what relationships are already collected. For example, many publishers collect funding information, but are not yet using grant IDs registered by funder members. Other considerations include which relationships may not be accommodated by their existing systems and what curation may be needed for some types. For example, how a new version is determined will vary among organizations.

Librarians and other **consumers and curators** of the metadata may also have limitations on integrating and using this information, particularly when there is not a critical mass of it, which is why the publisher role is such a big one in realizing the Research Nexus vision. Of course, besides ingesting relationships into their systems, metadata users can also contribute by using relationship information in research and analyses such as bibliometrics and by advocating for resources to make better use of these connections.

Crossref's role, like that of other **infrastructure providers**, is to make it easier to register and retrieve relationships and to make the benefits of those efforts clearer.

Community change like this is so often an iterative, chicken-and-egg process, and realizing the Research Nexus is unlikely to be an exception. For all stakeholders, feedback and collaboration play a vital role in advancing our collective efforts, for example on agreed-upon best practices.

6. The Research Nexus and Principles of Open Scholarly Infrastructure

On a related note, it is important to mention that the Research Nexus reflects the Principles of Open Scholarly Infrastructure (POSI) [3], a set of guidelines around governance and sustainability for scholarly infrastructure organizations and initiatives. The sustainability of organizations and initiatives that facilitate and maintain information, such as relationships in the metadata, is a key part of the persistence of the scholarly record. Both POSI and the Research Nexus are aspirational, focused on openness and broad stakeholder participation. Both recognize connections and interdependencies as critical to a healthy, robust research landscape.

About the Authors

Jennifer Kemp is Head of Partnerships at Crossref, where she works with members, service providers and metadata users to improve community participation, metadata, and discoverability. Prior to Crossref, she was most recently Senior Manager of Policy and External Relations, North America for Springer Nature. Her experience in scholarly publishing began with her work as a Publication Manager at HighWire Press, where she had a variety of clients publishing in a wide range of disciplines. Jennifer's perspective on the industry remains influenced by her years as a librarian and she is active in a number of community initiatives. At Crossref, she facilitates the Books Interest Group, Funder Advisory Group, and the Metadata User Working Group. She also serves on the Next Generation Library Publishing Advisory Board, the

Library Publishing Coalition Preservation Task Force, and the Open Access eBook Usage (OAeBU) Board of Trustees. E-mail: jkemp@crossref.org; <https://orcid.org/0000-0003-4086-3196>.

Patricia Feeney's role as Head of Metadata at Crossref was created in 2018 to bring together all aspects of metadata, such as their strategy and overall vision, review and introduction of new content types, best practice around inputs (Content Registration) as well as outputs (representations through Crossref's APIs), and consulting with the community about metadata. During her ten years at Crossref she has helped thousands of publishers understand how to record and distribute metadata for millions of scholarly items. She has also worked in various scholarly publishing roles and as a systems librarian and cataloger. E-mail: pfeeney@crossref.org; <https://orcid.org/0000-0002-4011-3590>.

References

- [1] L.M. Schriml, M. Chuvochina, N. Davies, E.A. Eloë-Fadrosh, R.D. Finn, P. Hugenholtz et al., COVID-19 pandemic reveals the peril of ignoring metadata standards, *Scientific Data [Internet]. Springer Science and Business Media LLC* 7(1) (2020), Available from: <http://dx.doi.org/10.1038/s41597-020-0524-5>, accessed September 1, 2022.
- [2] Metadata 20/20 [homepage on the Internet]. Stakeholders [cited 1 June 2022]. Available from: <https://metadata2020.org/learn-more/stakeholders/>, accessed September 1, 2022.
- [3] The Principles of Open Scholarly Infrastructure [homepage on the Internet]. [cited 1 June 2022]. Available from: <https://openscholarlyinfrastructure.org/>, accessed September 1, 2022.