Role of Crossref in journal publishing over the next decade

Ed Pentz

Crossref, Oxford, United Kingdom

Introduction

Global cooperation to improve scholarly research is critical, so it is great to have this opportunity to share information here. I do note that Crossref is very well represented and supported in Korea. We are delighted to have the Korean Council of science Editors as a Crossref member. Moreover, we have had the organization's support for many years. It is also significant that Kihong Kim has served on the Crossref board of directors since March 2021 [1]. In addition, Jae Hwa Chang has been a Crossref ambassador [2] since January 2018. I will be talking about Crossref's role in journal publishing and some of the things we are looking forward to over the next decade.

Overview of Crossref

First of all, I want to give a quick overview of Crossref, what we do, and our current status. Crossref is a not-for-profit membership organization with the goal of making research outputs easy to find, cite, link, assesses, and reuse. The leading service that Crossref provides is a registry of metadata for scholarly content to enable reference linking. The metadata includes persistent identifiers, also referred to as PIDs for short. The persistent identifiers we assign are digital object identifiers (DOIs). Crossref DOIs are citation identifiers for scholarly content including grants, preprints, articles, chapters, proceedings, standards, reports, protocols, dissertations, reviews, comments (conferences, video, blogs soon). The majority of our content is scholarly journals, although that is changing. However, one of the things that has been changing, and will be changing even more over the next few years, is moving beyond a focus on getting persistent identifiers, and recognizing the larger value of metadata, and services to disseminate and reuse the metadata.

Some publishers come to Crossref and think that getting a DOI is the most important thing, they need to do. They might see Crossref as a seller of DOIs. The DOI is important and part of the foundation of many of Crossref's services, but it is not the most important thing that Crossref provides. We now regularly emphasize that Crossref is an open foundational infrastructure [3,4]. This means that our metadata is made available in a persistent, sustainable, open way and that others build on it and incorporate it into their services. So over the next ten

Received: February 1, 2022

Accepted: February 8, 2022

Correspondence to Ed Pentz
epentz@crossref.org

ORCID

Ed Pentz
<https://orcid.org/0000-0002-5993-8592>

This essay was presented at the 10th Anniversary Symposium of the Korean Council of Science Editors through Zoom on September 8, 2021 (Korea standard time).

years, it is vital to have good quality metadata registered with Crossref. What is good quality metadata in the Crossref context? It is all about relationships and other identifiers in that metadata. This enables Crossref and others to connect many different content types and see how they relate to one another and the broader scholarly research ecosystem—we refer to this as the ‘research nexus.’ ‘Nexus’ is not a very common English word, but it is about a set of connections, a set of relationships.

PIDs and Metadata Connections and Services

To build the research nexus, Crossref is capturing the connections between authors, funding, funders, universities and research institutions, data, and publications—journals, journal articles, books, book chapters, conference proceedings articles, preprints, and other types of content. So I would like to show where things are now in terms of Crossref statistics. We have metadata for 126 million scholarly content items representing 10% growth from the middle of 2020 to 2021 (Table 1). As a result of the pandemic, we have seen an increase in the amount of content being published and registered with Crossref.

The number of journal articles has gone up 7.7% and that of books, 9.9%. A 64.8% increase to over 700,000 in preprints has been a big trend, and that trend is going to be continuing. In most cases, preprints are connected in the metadata to the

published version of record related to the preprint. Making sure we reliably capture the preprint to version of record links is something we are working on improving. We have ORCID IDs to identify authors uniquely. That has been growing quite a lot, and readers can see also that we are collecting references and making those references openly available, and now abstracts have increased. This all helps fill in the research nexus and capture the scholarly record.

For journals, it is about more and better metadata going forward and registering it with Crossref. Korea is one of fastest-growing countries in terms of Crossref membership, although Indonesia was the fastest in 2020. Crossref started out being very focused on North America and Europe. However, now it is global, with many members in Asia and many other parts of the world. I will talk about our strategic goals for the next few years.

New Public Roadmap

We happily put out our first public roadmap; it is a Trello board and available openly on our website (<https://bit.ly/crossref-roadmap>) (Fig. 1). We have many things in progress, and the roadmap covers what Crossref will be doing over the next year or two years if anyone wants to see more detail. But, I wanted to talk now about looking forward to over the next ten years. Two of the big things I want to talk about, are *principles of open scholarly infrastructure* and come back to the re-

Table 1. Growth rate of the registered identifiers in Crossref from June 2020 to June 2021

Content	As of June 2020	As of June 2021	% change
Total content items registered	114,895,544	126,512,688	10.1
No. of journals	79,771	92,347	15.8
No. of journal articles	82,272,390	88,596,671	7.7
No. of books	1,458,072	1,602,245	9.9
No. of book-related records	17,351,895	19,985,430	15.2
No. of conference proceedings	73,022	80,632	10.4
No. of conference papers	6,404,251	6,901,078	7.8
No. of preprints	429,210	707,546	65.0
No. of preprint-to-article links	95,231	141,827	48.9
No. of unique records with Funder IDs	4,025,618	5,295,718	32.0
Records with one or more authors with ORCID IDs	4,324,091	6,402,765	48.0
Total works auto-pushed to ORCID	3,863,327	6,695,714	73.0
Records with references	50,950,476	55,786,727	10.0
Members registering references	5,759	6,979	21.0
Members with open references	2,616	3,825	46.0
Records with abstracts	6,392,159	11,714,465	83.0
Members registering abstracts	5,543	10,737	94.0

Fig. 1. Screenshot of Crossref's Trello on the Crossref roadmap. Source: <https://bit.ly/crossref-roadmap>.

search nexus that I was talking about before. Going forward, it is all about untangling complexity. By working with the whole community and our member publishers, Crossref wants to help untangle some of this complexity and highlight critical issues for journals over the next ten years.

Key Issues for Journals from 2021 to 2031

The key issues for journals are trust, integrity, and transparency. Good quality metadata helps people make their own assessments about how trustworthy content is and enables journals to demonstrate how they ensure that content is high quality. In terms of transparency, for instance, publishing peer review reports demonstrates that peer review was done and what the outcome was. Metadata is critical going forward particularly in demonstrating that funder and institutional mandates have been adhere to. There are many exciting developments. Preprints have grown in importance—but, it is important the preprint metadata includes a link to the version of record—and the version of record needs to link back to the preprint. I haven't mentioned machine learning and artificial intelligence (AI), but these will have significant impacts as well. For example, detecting image duplication and manipulation is becoming more important.

Principles of Open Scholarly Infrastructure

Rather than talk about the technology changes, I am going to focus on a couple of key areas of the *principles of open scholarly infrastructure*. Crossref is moving beyond talking exclusively about persistent identifiers, which are important but are only part of the equation. Crossref has adopted a set of principles and best practices and is working with other organizations to adopt and generate a discussion about how open scholarly infrastructure organizations and projects should operate. There is a class of open foundational scholarly infrastructure, for example, Crossref, DataCite and ORCID, and we think it is essential to be clear about how we operate and what our principles are. Thus, we feel that these principles are taking us forward and important to making our vision become a reality within the next ten years. We want a rich and reusable network of relationships, captured in metadata, connecting research organizations, people, and actions. A scholarly record should be built on forever for the benefit of society. To be clear, Crossref is open and nonprofit, but many commercial services build on and incorporate Crossref metadata into their services. Commercial services are valuable but the open infrastructure should be non-profit. We need an open layer to enable many different services. Therefore, we want to capture more information about research activities and out-

puts, and organizations, and we want them to all have persistent identifiers and open, standardized metadata that is available through human and machine interfaces, and we think that the *principles of open scholarly infrastructure* help achieve this. There are 16 principles and best practices divided into a couple of different areas. We think it is essential that there are open, sustainable services that open research is based upon

Fig. 2. Seven organizations who have adopted the *principles of open scholarly infrastructure*: Crossref, DataCite, OpenCitations, OurResearch, Journal of Open Source Software Blog, Dryad, and ROR (Research Organization Registry).

Fig. 3. Research nexus that links a variety of formats of articles. Source: <https://www.crossref.org/blog/the-research-nexus-better-research-through-better-metadata/>.

Fig. 4. Crossref's metadata APIs. Source: <https://www.crossref.org/services/metadata-retrieval/>.

going forward. We have had seven organizations (Fig. 2) now who have adopted these principles, and we are hoping to have more, and we are in discussions with other organizations about it. It has been going well, and there will be more happening about this over the next few years.

Research Nexus

So I do also now want to come back to the research nexus. Metadata helps journals deal with critical issues such as data reproducibility, research and editorial integrity, reporting, and assessment. Journal articles are connected to lots of other things, and there are many other research outputs, and we want to link it all up. Video, data, software, preprints, and protocols all need to be identified; they are part of the scholarly record (Fig. 3). We are going to be working to help capture that information, and it does mean that journals and organizations that publish the journals need to have metadata and persistent identifiers. We are very interested in encouraging that and making sure this happens because, of course, it improves scholarly research and benefits society. We have this chart that shows the different aspects of the scholarly ecosystem (Fig. 4).

Crossref is working to go beyond basic bibliographic metadata. We have added funding and licensing information and this helps determine if something is open access or whether it meets a mandate from a funder or institution. We have full-text URLs for Text and Data Mining and offer Similarity Check, a service for screening manuscripts for possible plagiarism and we track updates, corrections and retractions through Crossmark. Most recently, we added grant identifiers and funders are joining Crossref to register grants with us. We also work with DataCite and the California Digital Library on ROR (Research Organization Registry), which provides persistent identifiers for scholarly organizations.

Conclusion

We are asking our members, our publishers, which includes

journals, to collect a lot of information, and we are developing tools to help that happen. Crossref was originally set up to make reference linking between journals easier but we have grown into an open, enabling infrastructure connecting many different stakeholders to enable open research. This will accelerate over the next 10 years. We want an integrated, efficient, sustainable, comprehensive open scholarly infrastructure. All research outputs and activities should be identified and connected with metadata expressing these relationships and making them available through the machine and human interfaces. The ultimate goal is to help researchers focus on their research and advance human knowledge.

Conflict of Interest

Ed Pentz has been an Executive Director of Crossref since 2000. No other potential conflict of interest relevant to this article was reported.

Funding

The authors received no financial support for this article.

References

1. Huh S. Presidential address: the Korean Council of Science Editors as a board member of Crossref from March 2021 to February 2024. *Sci Ed* 2021;8:1-3. <https://doi.org/10.6087/kcse.222>
2. Chang JH. Reflection on 4 years in the role of a Crossref ambassadors in Korea. *Sci Ed* 2022;9:69-73. <https://doi.org/10.6087/kcse.266>
3. Crossref. Strategic agenda [Internet]. Oxford: Crossref; 2021 [cited 2021 May 28]. Available from: <https://crossref.org/strategy>
4. Hendricks G. The road ahead: our strategy through 2025 [Internet]. Oxford: Crossref; 2021 [cited 2021 Jun 3]. Available from: <https://crossref.org/blog/the-road-ahead-our-strategy-through-2025/>