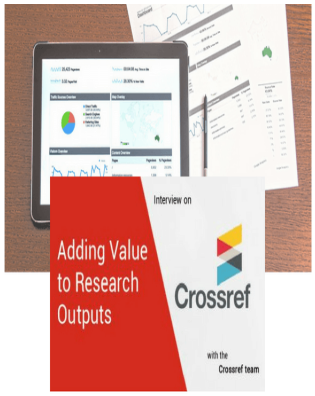

Description

Research data is growing exponentially and making all of this information easily accessible to researchers, publishers, and research institutions can surely be a daunting task. Crossref is an organization that aims to make scholarly communications better. They offer a number of services related to research metadata from various content types such as journal articles, preprints, books, and datasets.

As part of our interview series on Connecting Scholarly Publishing Experts and Researchers, we had the opportunity to speak with the Crossref team.

Could you give our readers a brief idea about Crossref?

We are a not-for-profit organization that helps to put scholarly content in context, however, we are much more than that—we are a sagacious configuration of staff, board, working groups, and committees as well as a broad range of collaborators, users, and supporters in the wider scholarly communications community. Everything we do at Crossref is designed to put research outputs in context so that the content our members publish can be found, cited, used, assessed, and re-used.

What are the primary areas of focus for Crossref in 2018?

Our focus for the coming year revolves around four key areas: improving metadata, expanding to help new constituencies, simplifying and enriching our services, and collaborating where we can with like-minded organizations. Everything we are planning for the coming year (and beyond) fits into one of these four strategic priorities. For example, we'll be looking at developing a persistent identifier (PID) for research grants and related activities as well as a central PID for organizations. We're also introducing dashboards where our members can see how they're doing in terms of metadata completeness, and overhauling our administration system so members can see their deposit queue and the status of content registration.

What are some of the recent achievements at Crossref?

The year 2017 was a great year for us—we were able to achieve a number of major milestones, such

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

as the Beta launch of our new and exciting service Event Data. In September, we were a big part of the collaborative launch of Metadata 2020, which aims to rally and support the community around the critical issue of sharing richer metadata for research communications. In November, we introduced a new service called Metadata Plus, which ensures our APIs continue to have genuine utility, and that platforms and tools like the library, search, analytics, and other systems can leverage our metadata to increase their offerings for their users, and the discoverability of members' content.

We are also very proud to report a record-breaking 1,939 new members in 2017—with the highest number of new members joining from Indonesia, Brazil, Japan, Turkey, and Russia.

Could you tell us some of the challenges faced by the Crossref team?

Our ever-present challenge is to continually find effective ways to communicate the wider story around the importance of open infrastructure and metadata—especially as our member base becomes more and more global. We have a strong Outreach team based across the US and the UK that very actively go out and speak at what we call Crossref LIVE locals—regional events that are local for the attendees. These events provide a much-needed face-to-face platform for us to share information with the research community on how to enrich their metadata deposits (and why they should), as well as covering the services Crossref offers that could help to facilitate this. Earlier this year, we also launched an Ambassador Program which will go a long way in helping us reach more members around the globe.

One specific recent challenge—within the context of obtaining richer metadata—has been conveying the options for members in how references are distributed through our services. Around one-third of our members' deposit reference lists from their journal articles, usually as part of their participation in our Cited-by service. Historically, references were not included in all our metadata delivery services and were subject to case-by-case opt-outs. However, in July 2017, the Crossref board voted to update the options for reference distribution. Members can choose for their references to be "public", "limited" or "closed" with the default being "limited" (only available to Metadata Plus users) and case-by-case opt-outs are not allowed. Members can inform us to set their references to 'open' and, while many have , there are still some publishers unaware of the options. References set to "open" are available through all our Metadata Delivery services, including the REST API and bulk data dumps, without restriction, to any interested party, an important contribution to the research endeavour as a whole.

Could you share some details on how Crossref is making citations more open through the "Initiative for Open Citations"?

Crossref is not affiliated with I4OC but they are a Metadata Plus user and can get references from us that are set to "open" and "limited". I4OC was set up by consumers of Crossref metadata and some publishers who are Crossref members to encourage more members to set their references to "open" to better benefit scholarly research and communications. We're happy to hear from any member who wants to set their references to "open" but it's a decision for each member to make – Crossref is neutral.

What is a DOI? How does it help authors?

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

A Digital Object Identifier (DOI) is an identifier system that provides unique alpha-numeric identifiers that can be turned into—and be used as—a URL. A DOI is applied to a particular piece of intellectual property, usually in an online environment. Unlike a URL, however, a DOI specifies the *content* of an online object, not the location, and is, therefore, a "persistent" identifier, as it remains associated with the object irrespective of changes in the object's web address. Crossref is just one of eleven members of the DOI Foundation and one of two worldwide ones (DataCite is the other with whom we work closely).

Simply put, a DOI can help an author's work become discoverable by uniquely identifying it, thereby providing a way to link to it long-term—so that it can be found, cited, shared, linked to, and ultimately used by others.

How is Crossref making peer reviews more citable and discoverable?

Publishers have been registering peer reviews with us for a while but historically these were embedded within other content such as an article, a dataset, or a component. To enable peer reviews to be deposited as a separate content type we extended our infrastructure to include metadata from the whole peer review history—referee reports, decision letters, author responses, community comments—and across all review rounds. In October last year, we announced that we're fully open for peer review deposits and as of the end of June 2018, we have around 12,000 registered items of peer review content.

What benefits does the Crossmark service offer to authors and journals?

Crossmark allows publishers to present "trust signals" to readers in a consistent way, so they can show the rigor that went into the work (and any additional information they choose). Crucially, it's a way to alert readers (even of saved PDFs years later) if there has been an update or a retraction. This is science doing its job of being self-correcting, so it is something all journals see as a crucial part of their job. Additionally, this update/correction/retraction information is available through our open APIs to be displayed on third-party tools like databases and reference manager systems.

Crossmark from Crossref on YouTube

What is Event Data? How does it help in understanding various interactions with online scholarly research?

Scholarly articles can be mentioned anywhere—for example, lots of people use Twitter to talk about research and we see headlines about research in the international newspapers all the time. Moreover, because we can't go to platforms like Twitter and the New York Times to ask them to register 'assertions', we built a system in collaboration with DataCite called Event Data (currently in Beta) to monitor platforms like Twitter, and then extract links from the mentions. We also, crucially, monitor each others' activity so that relationships between data and publications are recorded and revealed

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

through each service. Crossref Event Data monitors a number of platforms and brings research activity into one place, recording where research has been bookmarked, linked, recommended, shared, referenced, commented on etc., across the whole web—and beyond publisher platforms.

Joe Wass, Principal R&D Engineer at Crossref, has been an integral part of the Event Data story from the beginning and his blog posted earlier this year is a great read that provides a much deeper insight into Event Data including some of the challenges faced during its production.

What is the REST API? How does it help tie together all other Crossref services?

The REST API is our public machine interface that the community uses to query metadata records—about journal articles and preprints, books and book chapters, conference proceedings, standards, datasets, and component material—and it receives almost 200 million queries each month on our 98 million (and counting) metadata records. It's fully open and apart from the 'public' option, we encourage the use of the 'polite' route which just means identifying yourself by email in the query headers. The 'Plus' option mentioned above is the subscription option with all the extra services like a fast turnaround of queries, and it also includes access to our OAI-PMH API too if needed.

Every service that Crossref provides is based on our metadata, and our APIs expose all of that metadata. For example, you can query for Crossmark status updates (information on corrections and retractions) or all of the content that cites a particular funding agency. It's all there in one machine-readable place.

Over the past year or so we have been collecting use cases from members that actively utilize the Metadata APIs and we have turned these into a Metadata APIs blog series so that we can share their stories with the wider community.

Metadata APIs from Crossref on YouTube

How does Crossref cater to ESL countries like China, Korea, and Japan?

Our member base has been rapidly growing in its diversity, especially in Asia, with hundreds of new members from Asia in 2017 alone – so we have been actively creating ways to share information in languages other than English for some time now.

Earlier this year, we launched our service videos in many different languages—English, French, Spanish, Brazilian, Portuguese, Chinese, Korean, Japanese, and Arabic—with Indonesian and Russian in the plans for later this year.

I mentioned earlier that our recently-launched global initiative called the Ambassador Program helps us reach members that are further afield. Crossref Ambassadors are based around the world and help with content translation, running webinars, and managing training sessions on our services in their own local languages.

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

In the near future, we plan to launch an open community discussion forum where members (and others) will be able to talk to each other in their own languages—and we hope that later down the line forum members will discuss and share support issues and solutions between themselves and in their own local language.

Does Crossref adopt any specific measures to prevent the indexing of articles from predatory journals?

We have over 10,500 members residing in 120 different countries and more than 98 million content items have been registered with us by these members. We are very proud of our diverse, global membership and work hard to ensure that we are inclusive and that there are minimal barriers to participating in Crossref.

We have a streamlined membership application process in place where we ask for different types of information from potential members—and where we make it clear what being a Crossref member means and what the membership obligations are.

We can't, however, assess the quality of our members' content or verify our members' publication processes and procedures. It's not our role or part of our mission to do these things. There are many organizations and services which we support that help assess content and what goes into creating high-quality research outputs: COPE, DOAJ, Think.Check.Submit.

In your opinion, how is technology changing the landscape of academic publishing and what are some of its challenges?

More and more of our members seem to be opting for open source publishing tools like the Open Journal System from the Public Knowledge Project. They're also concerned with things like detecting potential plagiarism for conferences, papers, and even images. We don't have full solutions for this (yet) but there are possibilities to explore. They're also talking a lot about Artificial Intelligence and Machine Learning. We've been experimenting a little. But back to the open source trend – we're planning to open source as much as we can over the next few years.

What are some of the initiatives being taken to broaden the scope of partnership with publishers, research libraries, and organizations?

One of the very practical ways we are collaborating is simply by having research libraries and organizations join as members. We say "come one, come all" which means that if you post or publish something related to research—and commit to maintaining it—then you are eligible to join Crossref. We have data repositories, scholar-publishers, university libraries, government agencies, annotation tools, preprint services, and more. It's kind of the unknown membership trend — that for years now: we are not just publishers!

DataCite and ORCID represent other parts of the community and other related needs, and we partner quite specifically with them as fellow open foundational infrastructure services.

We're also working with research funders to develop identifiers for grants including awards and use of

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license

facilities. We're also involved with Metadata 2020 which is truly trying to break down the silos in our community, advocating for richer, connected, and reusable, open metadata, for the benefit of society.

Can you share some Crossref vital statistics that our readers might benefit from?

Yes, we currently hold over 98 million registered content records and expect to hit the 100 million mark sometime later this year. More than 68 million of these records have full-text links, and almost 3 million contain some kind of funding information—increasing the level of metadata records containing funder information is one of our focus areas for the coming year. An overview of these and other Crossref vital statistics is available on our website dashboard.

You can check out the Crossref website and also follow them on Twitter @CrossrefOrg

It was a great pleasure to talk to the Crossref team. We sincerely thank them for taking the time to be a part of this interview and also wish them all the very best in their future endeavors!

(This interview is a part of our interview series of Connecting Scholarly Publishing Experts and Researchers.)

Category

1. Expert Interviews
2. Industry News

Date Created
2018/07/17
Author
eneditor

Copyright: Enago Academy under the CC-BY-NC-SA 4.0 license