Scaling article/data linking infrastructure using the Scholix framework

Crossref adopts the Scholix Metadata Schema for Exchange of Scholarly Communication Links

The challenge

The Crossref challenge runs two ways. Our publisher members register metadata with us, and when that metadata cites related research data, whether as a reference or as the data underlying the article itself, we want to collect the information in a standardized format (with related identifiers). We also want to let publishers do so in a way that it enables us to let the repository hosting the data know that the data was cited.

From the publisher side, there's also interest in seeing which datasets reference publications - in some cases these links are bidirectional and in some cases they are not.

Being able to harvest both article/data links and data/article links in a standard format then helps this information to be used by consumers of bulk data-article links to build services or metrics and graph and link aggregators or hubs.

We didn't want to reinvent the wheel to communicate this information. Starting from scratch to define a metadata format, minimally required information etc. takes a lot of time and effort. Buy-in from the community takes even more. sing existing initiatives and recommendations is frequently our preferred method.

The RDA outputs adopted

We adopted the recommendation from the Scholarly Link Exchange (Scholix) WG. This group created the Scholix metadata specification for the expression of scholarly links between literature and datasets which is exactly what we were trying to achieve.

We work closely with our community, including publisher members and with the team at DataCite. They involved us in the Scholix Working Group so that we could provide input and implement information on article/data links in a way that would standardise the information across different stakeholders to maximise use of the information by any interested party. The data we provide and the recommendations adopted are openly available, both of which are important to our support of open scholarly infrastructure.

The impact of the adoption

The data we've been able to provide in the Scholix format is being used by tools like OpenAire's Scholexplorer service, which is now incorporating over 279,457,716 Scholix links. Not all those links come from Crossref, but that's the point; others providing links in the Scholix format all feed into a shared pool of information.

One of the main benefits of Crossref adopting the Scholix format is that it creates efficiencies for publishers in sharing their article to dataset links so that they can be used by the wider community in a standard way. When publishers send us links to data in their Crossref metadata in specific formats, we do the rest - making these links available via Event Data and therefore in the Scholix format, too. So, any publisher registering this information with us is automatically

Find out more at:

www.rd-alliance.org/recommendations-outputs

Visit   or write us at

rd-alliance.org | enquiries@rd-alliance.org

providing Scholix-compliant information, without having to do additional conversion/workflow tasks which can have time and therefore cost associated with them.

The easier it is for publishers to share this data, the more these links can be used to shore up connections between articles and data, help with reproducibility and reporting, and ensure that researchers get credit for sharing their data.

The adoption process

At Crossref, we have a service called Event Data which captures and records events that take place all over the web and makes the data openly available via an API. Two items related to the information of interest to the Scholarly Link Exchange WG that Event Data collects are:

» Links from the Crossref metadata to DataCite DOIs

» Links from DataCite metadata to Crossref DOIs

We worked to implement an endpoint to the Event Data API which makes the Event Data information related to the article to dataset links from Crossref and the dataset to article links from DataCite available using the Scholix schema.

The main resources we needed were initially technical, to spec-out the work and build the Scholix endpoint for Event Data. After that, we needed to spend time working on outreach activities to let our members know about this work and to encourage them to provide article/data links in the metadata they register with us to realise the benefits of the work we have done during the technical implementation.

It's an important piece of the puzzle for us, as we work with other groups to help connect research through better metadata.

Lessons learned

At Crossref we have experience in adopting standards, however this was the first RDA Recommendation we've adopted. The documentation was comprehensive for us to work with.

The Scholix Working Group also provided good support. It gave us a touch-point with other adopters and connected us with initiatives like Make Data Count and the STM Research Data Year which are helping us with the community aspect of driving the awareness and adoption aspects of article/data links, which are key to supporting the recommendations we've put in place.

About Crossref

Crossref makes research outputs easy to find, cite, link, assess, and reuse. We're a not-for-profit membership organization that exists to make scholarly communications better. We have over 12,000 members who were initially from the publisher community, but that range of stakeholders we work with has expanded a lot since we were founded in 2000, and encompasses funders, research institutions and thousands of metadata users. We have a store of metadata covering over 116 million content items, many of which are journal articles, but we also have books, conference proceedings, preprints and other items (it's expanding all the time).

Contact person:
Rachael Lammey
(Programme Manager at CrossRef)

Find out more at:

www.rd-alliance.org/recommendations-outputs

Visit   or write us at

rd-alliance.org   enquiries@rd-alliance.org