Digital Preservation of Scholarly Content at Crossref

Rachael Lammey
rlammey@crossref.org

Crossref (https://crossref.org) is a not-for-profit membership organisation, established in 2000 to provide persistent identifiers for scholarly content - journals, books, conference proceedings and more. By 2017, Crossref membership has expanded to over 7,000 members with over 600 based in Brazil.

Crossref members deposit metadata, in XML format, on the content they publisher, often using systems like Open Journal Systems (OJS) to do so. The metadata they provide contains valuable bibliographic information such as publication, title, authors (including ORCID iDs), reference lists, abstract, and the Digital Object Identifier (DOI) for the work.

It also contains the current location of the piece of content on the web, via a URL. Importantly, if the content moves to a new website, the publisher can come back to Crossref and update their metadata with us so that the DOI always points to the content in its current location. That way, any publisher or database linking to the content using that DOI will be automatically redirected to the current location of the content, so researchers can always find the related research quickly and easily.

Over the past number of years, Crossref has expanded the information it collects on publications. It is no longer just bibliographic metadata as other, supplementary information can also maximise the discovery and reuse of published research. For example, Crossref members can deposit license information in the form of a URL which then shows anyone what they are allowed to do with the content. They can also provide information on who funded the research so that this can be tracked by funders and publishers, details on if a paper has been corrected, retracted or updated in another way, and they can also show their archiving arrangements with organisations like CLOCKS and Portico.

Crossref collects all of this information and stores it in a database which can then be used by other parties - library systems, search interfaces and other discovery tools, publishers, other providers - to build tools and services around the research. The information from publishers is made available in a number of ways, and many users access it by using the Crossref REST API (https://api.crossref.org) which is free to use.

Because Crossref members are free to update the metadata associated with their publications as often as they want, Crossref's database is always updating so that the content can be found. The reason we hold metadata is also to do with preservation. Sometimes publishers or publications cease to exist and don't have archiving arrangements in place.

Even in a case like this, Crossref will keep the relevant metadata and it will remain available via Crossref so that even if the publication cannot be found, the information related to it will still be there (and may help in finding the paper some time in future). Portico preserves Crossref's bibliographic and reference metadata in the Portico archive so that this information will also persist long-term. However, because Crossref is a member organisation, funded by membership fees and fees for DOI deposits, it is sustainable as well, rather than being underpinned by short-term grants or loans. This is very important to ensure trust in the long-term preservation of scholarly metadata.

We would be keen to share this information with conference attendees to increase awareness of the importance of persistent identifiers for scholarly content. It is important that research remains easily discoverable and available so that it can be built upon by scholars both now and in the future.