science editing

pISSN 2288-8063
eISSN 2288-7474

Sci Ed 2014;1(1):13-18
http://dx.doi.org/10.6087/kcse.2014.1.13

Review

# CrossRef developments and initiatives: an update on services for the scholarly publishing community from CrossRef

Rachael Lammey

CrossRef, Oxford, United Kingdom

## Abstract

CrossRef (http://www.crossref.org) is a not-for-profit membership association of publishers. Since its founding in 2000, CrossRef has provided reference linking services for over 62 million scholarly content items, including journal articles, books and book chapters, conference proceedings, reference entries, technical reports, standards, and data sets. CrossRef also provides additional collaborative services designed to improve trust in the scholarly communications process, including Cited-By linking, CrossCheck plagiarism screening, CrossMark update identification, and the FundRef funder identification service. All of these services continue to develop to try to meet the evolving needs of publishers, and this article attempts to give an overview of them and highlight important updates which have taken place over the last 12 months.

## Keywords

Digital object identifier; CrossCheck; CrossMark; CrossRef; FundRef

**Received:** October 24, 2013
**Accepted:** October 31, 2013

**Correspondence to** Rachael Lammey
rlammey@crossref.org

**ORCID**
Rachael Lammey
http://orcid.org/0000-0001-5800-1434

This paper was given in the form of a presentation at the EASE/ISMTE joint meeting in Blankenberge, Belgium on 24th September 2013, but is being expanded for publication as a review article.

## Introduction

CrossRef is a not-for-profit membership organisation for publishers, founded in 2000. Those who work in the publishing industry will predominantly have heard of CrossRef in terms of the first service it offered, reference linking using the Digital Object Identifier (DOI). A variety of services provided or that will be provided by CrossRef inclusive of the DOI are presented in this paper. It aims to give hints for publishers, editors, or researchers to help them understand the meaning of present and new services by CrossRef.

## CrossRef Digital Object Identifier

With the advent of online publishing, publishers, librarians and researchers were finding that

links to published content from reference lists frequently failed as content moved around the web. They needed a way to create permanent links to content so that researchers would be able to link consistently to cited material, and the DOI is the method by which to do this. The DOI is to all intents and purposes a meaningless number, but it allows a piece of content to be located on the web. It works like this: publishers use CrossRef DOIs to link to content, usually from the references at the end of articles. Users click on those DOI-based links and are referred via the CrossRef database to the cited article at its correct location on the web. If content moves the publisher only has to update the CrossRef database once, and all of the publishers that are linking to their content using CrossRef DOIs will be redirected to the content in its new location (Fig. 1).

CrossRef is not the only organisation that provides DOIs. DOI is a trademark of the International DOI Foundation (IDF, http://www.doi.org) that appoints registration agencies like CrossRef. As such, many types of content have DOIs, including consumer movies, scholarly articles in languages other than English, reference works and data. Not all of these DOIs will have been issued through CrossRef and other agencies like DataCite, mEDRA, Movie Labs and CNKI can assign DOIs to content as well. Geoffrey Bilder's recent blog-post 'DOIs unambiguously and persistently identify published, trustworthy, citable online scholarly literature. Right? [1]' is an interesting expansion on this point and is useful to bear in mind that not all DOIs are CrossRef DOIs.

The reason that CrossRef DOIs are discussed in this paper is that the CrossRef DOI and associated metadata still forms CrossRef's core service. However, in the past 8 years CrossRef has offered an expanding number of services for its members based on this, due to demand from the scholarly publishing industry.

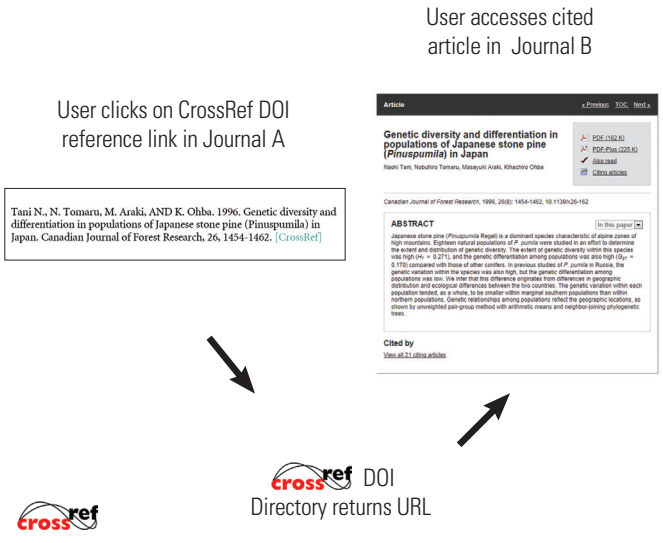


**Fig. 1.** Simple example of the CrossRef referral process. DOI, Digital Object Identifier.

## FundRef

CrossRef's most recently-launched service is FundRef (http://www.crossref.org/fundref) which is a standard way of reporting funding sources for published scholarly research. Because of the different ways that funding information is deposited by authors, and displayed and deposited by publishers, problems were arising. Funding bodies were not able to track the published output of funding as it was spread across so many publications and different publishers. Equally, because there was no standard list or format to collect this information from authors, publishers could not easily report which articles result from research supported by specific funders or grants. Institutions could not easily link funding received to published research, and there was a lack of standard metadata for funding sources making it difficult to analyze or data-mine.

FundRef is a collaborative solution to this problem, devised by and for the benefit of both publishers and funders. Both parties have an interest in the outcomes of FundRef, and both have well-established processes, one for recording the distribution of funds and monitoring the research process, and the other for ingesting, processing and publishing the outcomes of the research. The piece that has been missing is the one that links these two sets of processes, and that is where FundRef comes in, recording this link and making it more visible. CrossRef has just completed a year-long FundRef pilot that ran until March 2013, and involved American Psychological Association (APA), Wiley, the American Institute of Physics (AIP), Nature Publishing Group, Oxford University Press (OUP), IEEE and Elsevier, and funding bodies like NASA, the NSF, the Welcome Trust and the US Department of Energy. On successful completion of the pilot project the CrossRef Board approved the FundRef service to go into production, which happened on May 28th 2013.

One of the key things that came out of the pilot and is central to the project is an agreed taxonomy of funding bodies. The FundRef Registry has been created from a list donated to the project by Elsevier, and currently consists of around 4000 international funder names. The list data is and will be freely available under a CC0 license waiver. The Registry will be updated monthly, and new organizations suggested by publishers or funding bodies themselves will be added after curation. This is the list that publishers will use to collect information from authors on submission.

To explain in more detail how the process works, CrossRef hosts a funder registry which provides standard funder names

to publisher submission systems. Publishers ask authors, at submission, to provide the name(s) of the funding bodies and accompanying grant numbers. This funding information goes into publishers' production systems where it is stored as tagged XML and submitted to CrossRef with all of the other deposited metadata for each piece of content. Once the funding information is in the CrossRef database it becomes a searchable, either through our search interfaces or via one of CrossRef's application programming interface (API). Publishers, funders, and other interested parties can then query on a funding organisation or grant number to discover the resultant publications, or can look up a piece of content using other metadata and find out the funding sources.

Publishers will be able to display this funding information in a structured way. For those publishers who are participating in CrossMark, the funding data will automatically appear in the Record tab of the CrossMark dialogue box. CrossRef encourages publishers submitting FundRef information to also participate in CrossMark, as this further standardizes the location of the information for readers, but of course it can also be displayed on the publisher's site in metadata and full text (Fig. 2). The key piece is that the funding information is now centrally stored in the CrossRef database and can be queried. These three pieces of information of the DOI, the funding source(s) and award numbers are tied together in the metadata, making each of them discoverable via any of the other. Taking this a step further, once this information is in the CrossRef database and ORCIDs are also being deposited, you have a scenario in which you can look up a researcher, find their publications, and see how their research was funded, or look up a grant number, see its associated DOIs and

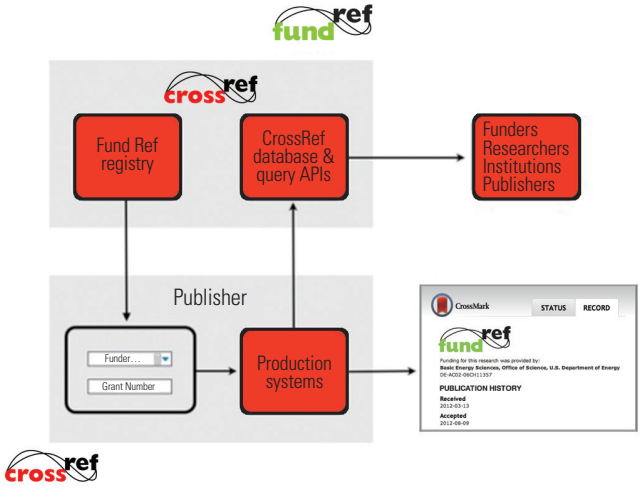


**Fig. 2.** The FundRef workflow. API, application programming interface (Reproduced from: http://www.crossref.org/08downloads/handouts/FundRef_Workflow.pdf).

which researchers contributed to those publications. So the information that the FundRef service collects can become a key piece of article metadata and aid discoverability.

## Text and Data Mining

Another initiative is a service related to Text and Data Mining (TDM), which is in the pilot stage at CrossRef. Researchers are increasingly interested in text and data mining published scholarly content, and this poses technical and logistical problems for scholarly researchers and publishers alike. Researchers find it impractical to negotiate multiple bilateral agreements with subscription-based publishers in order to get authorisation to TDM subscribed content, and publishers face similar problems in having to negotiate with multiple researchers and institutions who want to mine their content. All parties would benefit from support of standard APIs and data representations in order to enable TDM across both open access and subscription-based publishers. The service proposed by CrossRef would provide two major components to address the issue of text and data mining the scholarly literature:

- A common API that can be used by researchers to access the full text of content identified by CrossRef DOIs across publisher sites and regardless of their business model.
- A mechanism that can (optionally) be used by researchers and publishers as an efficient mechanism to provide "clickthrough" agreement of proprietary TDM licenses.

Both components would be free to use by researchers and the public. At present, this service is in the pilot stage, and a working group are having their technical teams review the proposed standards and sample code and thus make recommendations based on their experience implementing, using, and maintaining such systems. Some members of the working group will be examining the system exclusively from the point of view of publishers. Other working group members are working with researchers to evaluate the system. At the CrossRef Board Meeting in November 2013, a decision will be taken on if/how to launch CrossRef TDM as a production service. More information is available for publishers and researchers at: http://prospectsupport.labs.crossref.org/.

## CrossCheck

Whereas the TDM and FundRef services from CrossRef are relatively new offerings, (TDM being at the pilot stage as of October 2013), CrossCheck, powered by iThenticate, was launched in 2008 and continues to expand and develop. CrossCheck consists of two aspects to allow publishers to check their content for originality; a full-text database of scholarly articles, books and conference proceedings from

CrossCheck member publishers, and the iThenticate tool which is used to compare submitted documents against the database. The iThenticate system does not detect plagiarism, but it does highlight matches in text between the article that has been uploaded by the publisher and the CrossCheck database so that editors and journal staff can look at the overlap and determine whether it requires further investigation.

The last 18 months has seen a number of important developments regarding the CrossCheck service. Firstly, over 500 publishers have now joined CrossCheck and enabled the indexing of their full-text content in the database. There are now 38 million content items indexed for members to check their papers against, and CrossRef has seen a marked increase in the number of papers members are uploading to check for originality. In the month of August 2013, over 100,000 papers were submitted to iThenticate, and the continued growth in usage shows that publishers are taking steps to integrate CrossCheck into their peer review processes and try to ensure the originality of the content they publish.

Other CrossCheck developments in the past 12 months include improvements to the iThenticate system itself. The main change that users will have seen is the release of the Document Viewer (DV), which displays the fully-formatted document as it was uploaded rather than just the text itself. The DV aims to make the similarity reports easier to interpret by enabling users to see, more clearly, where the matching text sits within the document in order to put it in context. The size of files that users can upload has also increased from 20 MB to 40 MB and small match exclusion has been introduced. To explain how small match exclusion works, if it is set to exclude match instances below 10 words, every match the iThenticate system finds that is less than 10 words will automatically be excluded from a report. This provides users with the ability to customize report match sizes to help focus on more relevant potential misconduct issues and remove trivial matches from reports.

September 2013 saw an iThenticate release that enabled section exclusion functionality. CrossCheck users were reporting that they were seeing large matches because of text that was overlapping in the abstract or materials and methods section. Often these sections contain standardised wording or set phrases which match set text or phrases from other published papers. Using section exclusion means that users can decide to exclude the abstract and materials and methods sections from their similarity reports and just focus on the main body of the text.

CrossRef and iParadigms are working together to collect feedback from CrossCheck members in order to continue this schedule of developments and keep making valuable improvements to the service. Suggestions from members include adding the capacity to make the largest matches to a single source and the largest match in terms of number of words more visible in the system, and the ability to match equations, tables and figures. CrossRef will survey CrossCheck members in late 2013 to get more input on the development process, and CrossCheck User Groups have been run at the COPE European Seminar in London, the Council of Science Editors Annual Meeting in 2012 and 2013 and the CrossRef Workshops day at the CrossRef Annual Meeting. These will continue in 2014 and are proving a useful resource for users and CrossRef and iParadigms staff.

## CrossMark

CrossMark is also an optional service for CrossRef members and it went live in April 2012. At its simplest, CrossMark is a logo that publishers can place on their HTML and PDF content that identifies that piece of content as being maintained by the publisher. If a researcher clicks on the CrossMark logo, a dialogue box will appear which will tell them; whether there have been any updates to the content (i.e., if it has been corrected, retracted or supplemented in some way), if this instance of the work is being maintained by the publisher, where the publisher-version is and other important publication record information (if provided).

A service like CrossMark is important for researchers for a number of reasons. The first concerns PDF files. If a researcher has downloaded the PDF of an article from a publisher website, or maybe had it sent to them via email or accessed it from a document repository, they will have no way of knowing whether or not the article is current or if it has been updated unless they go back to the publisher website to check each time they access it. The other issue is that even on publisher websites, updates are displayed in different ways and in different places so it can be difficult to find this information, and it often doesn't carry through to third-party sites where content maybe hosted. All of these factors mean that researchers may run the risk of using or citing material that may have been corrected or retracted in their articles.

With CrossMark, all researchers have to do is click on the CrossMark logo, and this will tell them if the piece of content they're accessing is up-to-date, and if there are updates, it will link them, using the DOI, to a description of how the paper has been altered, alerting them to information they may otherwise have missed. To demonstrate the functionality, the CrossMark logo is displayed in one of the following forms on the publisher HTML and PDF content (Fig. 3). If a user clicks on the logo, they will see a dialogue box like this if the content is current (Fig. 4) or one like this to indicate that the article has been updated, in this case with a retraction notice (Fig. 5).


**Fig. 3.** CrossMark example (greyscale and black and white versions are also available).

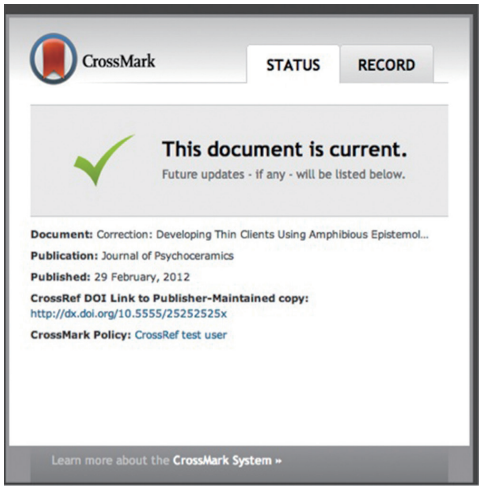


**Fig. 4.** CrossMark example for content with no updates.

CrossMark provides a simple, consistent way for researchers to access this information. It is also optional for publishers to provide additional publication record information alongside their CrossMark deposits. If they choose to do so, this information will be displayed on the CrossMark Record tab, which is at the top right of the dialogue box shown in Figs. 4 and 5. CrossRef does not dictate what information publishers can deposit, but so far publishers have been using it to communicate about publication dates, CrossCheck screening, supplementary materials, copyright and licensing information and funding information (sometimes as part of FundRef) (Fig. 6). Since CrossMark was launched, CrossRef has seen over 200,000 CrossMark deposits from a wide range of publishers including Elsevier, the Royal Society, F1000 Research, the International Union of Crystallography and a number of publishers from Korea including the Journal of Educational Evaluation for Health Professions.

CrossMark has also been integrated into some third-party tools such as Microsoft Academic Search who are displaying the CrossMark logo on relevant content within their index. Inera's eXstyles product is also supporting CrossMark. If eXtyles is being used for references, it will now provide a warning if a reference has a CrossMark record that indicates it has been "retracted," "withdrawn," or "removed" so that a researcher can exercise caution when citing that article. Cross-

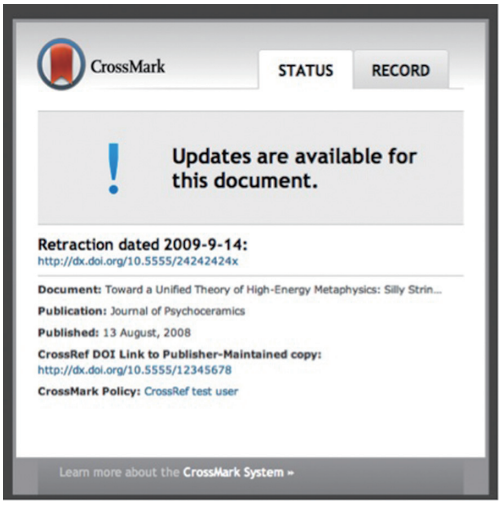


**Fig. 5.** CrossMark example for content that has been retracted.

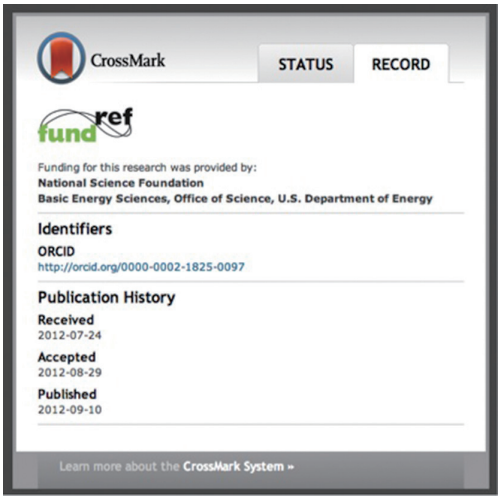


**Fig. 6.** Example of FundRef and other publication record information being displayed on the CrossMark record tab.

Ref expects CrossMark uptake to continue to grow, especially as it starts to be implemented by some of the major platform providers and publishers start to use it as part of their standard processes.

## CrossRef Labs Project

The final service to highlight is a CrossRef Labs project that will soon go live. CrossRef Metadata Search (http://search.labs.crossref.org) allows publishers, libraries and researchers to search across nearly 50 million CrossRef Metadata records for journal articles and conference proceedings. It supports features like ORCID, faceted searches, copying of search results as formatted citations, COinS (so that users can easily import results into Zotero and other document management

tools), an API so that users can integrate CrossRef Metadata Search into their own applications and basic OpenSearch support so that CrossRef Metadata Search can be added to a browser search bar. It is a simple way to search for a particular CrossRef DOI or CrossRef ShortDOI, or search for articles in a particular journal via the journal's ISSN, and it also shows funder information (if available) and links to any patents that cite a particular CrossRef DOI. CrossRef Metadata Search has been well-received thus far and provides a simple way for anyone to search the CrossRef metadata, and CrossRef members can expect more information on this piece of functionality when it moves into being a full production service.

## Conclusion

Since it was founded in 2000, CrossRef has aimed to support the scholarly communications industry by providing collaborative services for academic publishers. Starting with it's position as the official DOI link registration agency for scholarly and professional publications and providing reference-linking using the DOI, CrossRef has expanded the services that member publishers can make use of into areas like originality screening through CrossCheck, update identification through CrossMark, providing a standard mechanism to collect funder information via FundRef and is continually looking to provide additional services to support the industry. A text and data mining initiative and useful developments like CrossRef Metadata Search serve to try to continue to meet member needs as the academic publishing industry continues to evolve.

## Conflict of Interest

No potential conflict of interest relevant to this article was reported

## Reference

1. Crosstech. DOIs unambiguously and persistently identify published, trustworthy, citable online scholarly literature. Right? [Internet]. Crosstech; 2013 [cited 2013 Oct 23]. Available from: http://crosstech.crossref.org/2013/09/dois-unambiguously-and-persistently-identify-published-trustworthy-citable-online-scholarly-literature-right.html