ALPSP Advice Note No. 37

CrossRef

by

Amy Brand,
Director of Business and Product Development
CrossRef

CrossRef

What is CrossRef?

CrossRef's core service connects users to scholarly and professional publications by providing links from the references at the end of a publication to the cited items. This service functions on a cross-publisher basis, using a technology that also provides for link persistence. Numerous surveys have shown that cross-publisher linking is one of the features of online journals that users value most highly.

Cross-publisher linking offers several advantages to publishers. First and foremost, it brings readers to your publications and your website. By allowing readers to connect to your content from outside resources and locations, you will not only serve your subscribers better, but also create opportunities for much wider access to your publications. When a publisher registers content in the *CrossRef* database, thousands of participating organizations automatically pick up links to that content.

CrossRef can also be described as a *technology enabler* for publishers, because it works on services that publishers can create more efficiently collectively than they can on their own. In addition to reference linking, *CrossRef* also provides cited-by linking and metadata distribution services on a cross-publisher basis. *CrossRef's* success today is perhaps best measured by its impact on the research experience; researchers currently use *CrossRef* DOIs at a rate of about 12 million clicks per month.

CrossRef was launched in 2000, initially focusing on reference linking within STM journals published by different publishers. The technology used was the Digital Object Identifier, or DOI. *CrossRe*f has grown rapidly to include the broad spectrum of scholarly and professional fields, publishers, and content types, along with a range of services for publisher interoperability.

By November 2006, *CrossRef* had registered over 23 million scholarly items, from hundreds of participating publishers. *CrossRef* adds an average of 12,000 new records to its database each day, covering journal articles, books, chapters, conference proceedings, images, figures, datasets, working papers, and dissertations. Archival material is also included; the oldest content dates from 1769 - Volume 1 of the *Transactions of the American Philosophical Society*.

The DOI is a string of letters and numbers that both uniquely identifies electronic content, and provides a stable, permanent link to that content's location on the web. A DOI stays the same, even if there are changes in

ownership or location, because it's just a name used to look up an address in an updateable directory. This avoids the common problem of broken links associated with URLs.

For electronic books such as reference works, *CrossRef* provides a way of creating internal linking of components and references as well. Assigning DOIs rather than URLs to book chapters makes it easier to re-purpose content - for example, to create derivative works such as course-packs, which use a subset of the original components or rearrange their order - and to support 'pay per view' sales of individual chapters.

Benefits to Publishers

In addition to providing permanent links between content published by different publishers, *CrossRef* also drives traffic to publishers' sites. Once your publications are included in the *CrossRef* database, thousands of participating organizations – publishers, intermediaries and libraries – can automatically pick up links to your content. If your content is not registered in *CrossRef*, it is not 'visible' for linking in this way

Another advantage is that one agreement with *CrossRef* is, in effect, an agreement to link to and from content with all other *CrossRef* participants. A linking network on such a large scale would be impossible without CrossRef. In particular, the many smaller publishers that *CrossRef* now includes would have found it difficult or impossible to participate in widespread interlinking.

Publishers also benefit from being part of a collaborative platform for ongoing development of shared technologies, while maintaining control over their own business practices and how their content is accessed. Through inclusion of smaller publishers who might otherwise have been left behind in the move to interlink between information providers, *CrossRef* extends the network of accessible scholarship online.

Thus, CrossRef's three key benefits to publishers are:

1. a technology for permanent linking and content discovery
2. a convenient way of linking to and from many other publishers
3. shared development of collaborative technologies for the future.

What Publishers Have to Do

To participate, you have to become a *CrossRef* member. This entails submitting a membership application and agreement. These forms are available upon request from *CrossRef*. Membership is available to virtually any publisher of original scholarly material. In addition to journal and book

registration, *CrossRef* also supports registration of working papers, technical reports, dissertations, standards, and datasets. Members are assessed an annual fee, which is tiered according to gross publishing revenue.

Publishers use the core *CrossRef* system in two ways:

1. They deposit DOIs and metadata to enable inbound linking to their content.
2. They create outbound links from their references by querying the CrossRef metadata database to retrieve DOIs.

Members have an obligation to create outbound reference links, not just to enable inbound ones. This means that there is reciprocal linking to and from other member publishers. It also adds valuable functionality – permanent links that take you right to the appropriate content – to your own online publications.

Those publishers who have XML-based production workflows can deposit XML-formatted metadata records into the *CrossRef* metadata database, and query the database to find DOIs for their references, in large batches. For smaller publishers who may not have adopted XML, there are non-XML alternatives – a web deposit form for manual DOI registration, and a Simple Text Query interface for retrieving DOIs. The Simple-Text Query service parses unstructured, simple-text references into XML and returns any matching DOIs for those references. A simple cut-and-paste form accepts references formatted in a wide range of bibliographic styles. For editorial purposes, a publisher can easily check the accuracy of a reference by clicking on the returned DOI and viewing the bibliographic information available directly from the publisher.

Each deposited record must include minimal metadata (i.e. bibliographic information): a DOI, and a current URL. For a journal article, the descriptive metadata includes journal title, ISSN, first author, year, volume, issue, and page number. After a metadata record is deposited, *CrossRef* registers each DOI-URL pair in the central directory. When a user clicks on a DOI, it is the central DOI directory that matches it to the currently applicable URL. When a URL changes (for example, because a publication has changed hands, or a publisher has moved its content to a new platform), the content owner simply updates the appropriate once with *CrossRef*.

When publishers deposit records, they immediately enable live links to their content because other publishers, librarians, and intermediaries have processes in place to retrieve DOIs from *CrossRef*. What happens is that the publisher, library or intermediary submits a list of journal references or bibliographic records either to an automated XML query process or, if they don't use XML, to the Simple Text Query interface. This looks for matches in

the *CrossRef* database, and returns DOIs where matches are found. The returned DOIs allow a publisher to add permanent outbound hyperlinks to items already registered in the *CrossRef* system.

Linking does not assume that the full text is freely available to all; *CrossRef* is access-model neutral. When metadata and DOIs are deposited with *CrossRef*, publishers must have live response pages in place at the article level, to which incoming links will go. At a minimum this would consist of a full bibliographic citation, plus a way to gain access to the full text. The business model(s) for full-text access remain under the publisher's control. What most publishers do is to take users to the abstract page as a default, while permitting IP-authenticated users to go directly to the full text. Many publishers also offer pay-per-view options to non-subscribers. If the full text is available at no charge, as it is for a growing amount of content registered with *CrossRef*, then of course all users can view it immediately.

CrossRef in Libraries

Many researchers access content through the institutions with which they are affiliated. Because DOIs are assigned by publishers, they link to the version on the publisher's site. However, this is not always appropriate for users working in an institutional context. For instance, the institution may not subscribe directly to an e-journal, but may still be able to offer the user access to the target article through an aggregated database or through its print holdings. In addition, the library may wish to provide a range of navigational options beyond what is available at the publisher's own website.

In order that their products integrate with library linking systems, publishers are being asked to implement the OpenURL. People sometimes think of the OpenURL and the DOI as competing technologies, but they are not!

The OpenURL is a way of transporting metadata and identifiers describing a publication, to enable context-sensitive linking. A 'link resolver' (a system at the library site that can interpret incoming OpenURLs, take the local holdings and access privileges of that institution into account, and display links to appropriate resources) allows the library to configure its own links and services, including links to the full text, to a local catalogue to check print holdings, to document delivery or ILL services, to databases, to search engines, etc.

The DOI and the OpenURL work together in several ways. First, the DOI directory itself is OpenURL-enabled. This means that it can recognize a user with access to a local resolver. When such a user clicks on a DOI, the *CrossRef* system redirects that DOI back to the user's local resolver. It also allows the DOI to be used to extract metadata from the *CrossRef* database, in order for

the library to create the article-level OpenURL link. This means that when the institutional user clicks on a DOI, he or she is directed to the appropriate version of the item.

Thus, by using the *CrossRef* DOI system to identify their content, publishers in effect enable the use of OpenURL to access their products. Many publishers now use the DOI as the primary linking mechanism to their full text. Link resolvers can use the *CrossRef* system to retrieve the DOI, if it is not already available in the reference.

Recent Developments

The *CrossRef* network has expanded not only in terms of content coverage – to different content types and levels of granularity – but also in terms of functionality. Three recent developments are described here.

Forward linking is a way of tracking citations to a given publication as they occur. In addition to using *CrossRef* to create outbound links from their references, *CrossRef* member publishers can therefore now retrieve 'cited-by' links – links to other articles that cite their content. An optional tool is offered to *CrossRef* members, enabling them to display 'cited-by' links in the primary content that they publish. This is a natural extension of the *CrossRef* linking network, and will provide a better online reading environment for researchers and scholars.

As part of the same functionality, *CrossRef* also offers a 'match-alert' feature which saves users from having to query *CrossRef* repeatedly to find DOIs for citations that do not initially return a match. If the publisher chooses to enable this, then the system automatically sends an email containing the matched results when the relevant content is registered in *CrossRef*.

Multiple resolution enables publishers to present a menu of options to a user who clicks on a single link. These include the option to go to alternative sites for the same content or to different formats of a work (e.g. a print or PDF version); to view related resources; to drill up or drill down within the publication; to access associated metadata; to get more information about the author; and to purchase or acquire rights to the content. This can be implemented either as a pop-up menu or via an intermediate page hosted by *CrossRef*.

CrossRef Web Services is an optional service for publishers who would like *CrossRef* to disseminate metadata on their behalf. This can be useful to streamline the web crawling carried out by search engines and others, in order to optimize the functionality of their indexes. It consists of a suite of tools to enable authorized partners to collect metadata from multiple publishers, using a variety of metadata gathering implementations such as OAI-PMH. Standard

terms of use are provided for search engines, libraries, and other partners to use the metadata available from *CrossRef* Web Services.

How do I join?

If you are a publisher of original scholarly content and wish to join *CrossRef*, please email inquire@crossref.org and we will send you the appropriate membership forms and information.

January 2007