CrossRef and DOIs: New Developments

by ED PENTZ

INTRODUCTION

Profound changes to scholarly communications have been taking place since the advent of the World Wide Web and changes are continuing. Publishers and librarians are facing a new generation of students, researchers and scientists. The Pew Internet and American Life Project The Internet goes to College (Jones, 2002) found that 73% of students use the Internet more than the library for information searching. The Survey also revealed a danger for publishers of scholarly journals - that of losing readers. To quote from the survey:

"*Many students are likely to use information found on search engines and various Web sites as research material … and faculty often report concerns about the number of URLs included in research paper bibliographies and the decrease in citations from traditional scholarly sources.*"

In order to counter these trends, publishers and librarians must work together to make it easy for users to access authoritative scholarly sources online. This can be difficult to do since users' expectations of online journals have increased dramatically over the last 8 years and continue to increase. For most users these days content must be online and it must be linked. If it's not linked, it doesn't exist. Linking on the web works very well in conjunction with traditional references in journals. The practice of citing other articles has been around since the first journals appeared in the 17th century and it's a crucial part of the scientific process. So naturally, with online journals there is immense value in users being able to go to a cited item with one or two clicks.

Another important trend is the development of the article economy - journal issues are becoming less important. Most publishers now post articles online, "ahead of print", without volume, issue and page number, and for many journals the e-article is the article "of record". Because of this trend and the need for reference links, uniquely identifying each article, creating standardized metadata and creating a system to link references are of critical importance for publishers.

UNIQUE IDENTIFICATION OF CONTENT ONLINE

The Digital Object Identifier (DOI) is the answer to uniquely identifying content online. The DOI has been called a bar code for digital objects. It's a string of letters and numbers that uniquely identifies a piece of content and provides a persistent link to the location of the content. Broken and outdated URLs are a real problem on the web and DOIs help to solve this problem. DOIs are flexible and can be assigned to any type of content at any level of granularity - for example a DOI can be assigned to an article and to a figure, image or table from an article. Or looking at larger groupings, a DOI can be assigned to a book or a journal at the title level. The basic rule is that you identify whatever you need to identify. A DOI number is registered, along with a URL, in the central DOI system. By sending a DOI to the DOI Directory the user is linked to the URL registered by the publisher. The DOI Directory comes into play when a user clicks on a DOI link - the DOI system automatically resolves the DOI to the URL deposited by the publishers. This happens as an automatic redirect in the user's browser . Adding this level of indirection through a central DOI Directory helps ensure that DOIs are persistent. While the location of content may change, or ownership of content might change, the DOI itself does not change. Of course, publishers have an obligation to keep the URL up-to-date. The DOI supplements the traditional bibliographic data, it doesn't replace it.

HOW DOES CROSSREF FIT INTO THIS?

CrossRef is a non-profit membership association of publishers. By creating a non-profit membership organization, with by-laws and rules of fair play, the publishers overcame most of their competitive fears by creating a level playing field. CrossRef provides both a technical and a business infrastructure to enable linking and also provides a structure for determining policies and resolving issues. And, importantly, CrossRef establishes a central source for linking without the need for publishers to negotiate bi-lateral agreements and processes with other publishers. The CrossRef mission statement is: "To provide services that bring the scholar to authoritative primary content, focusing on services that are best achieved through collective agreement by publishers". CrossRef is also an official IDF (International DOI Foundation) Registration Agency (RA). This means that CrossRef is officially authorized to register DOIs for its members. The IDF runs the global DOI system and sets the basic standards for how DOIs work and what metadata is needed. As an RA, CrossRef uses the infrastructure provided by the IDF but also actively works with the IDF and other RAs on developing the DOI system. CrossRef sets the specific rules for DOIs and metadata for the scholarly community, collects the metadata and DOIs and represents its members on the IDF board, Technical Working Group and the Registration Agencies Working Group. CrossRef also works with, and establishes, guidelines and standards. For instance, CrossRef has many rules to insure that linking is reciprocal and fair across all its member publishers. Also, this year we published DOI Guidelines on the use of DOIs in journal articles and citations.

How Does CrossRef Work?

The first step is for publishers to deposit their article metadata, including a DOI and URL with CrossRef in a central metadata database. The publishers do not send abstracts and full text to CrossRef, just bibliographic data. Once all the article metadata is collected centrally, organizations (all types of publishers, vendors and libraries) can query the database using bibliographic metadata to find the DOI for an article that they want to link to. The metadata database acts like a telephone directory for journal articles. Once a DOI is retrieved from CrossRef, the publisher, database producer or library then adds a DOI link (it may say "CrossRef" or just "Full Text"). An end user then clicks on this link and the DOI resolves to the URL registered by the publisher. The user goes to the publisher's site and the terms of access to the full text are determined by the publisher - in most cases the users gets at least the abstract. CrossRef rules require that all users see at least a full bibliographic citation for an article. Subscribers to a journal will get the full text or the full text article can be offered at no charge. CrossRef is "business model neutral" and we are happy to have such publishers as PLoS (Public Library of Science) and BioMed Central as members.

Is CrossRef Working?

The answer is a resounding yes. There are currently 252 member publishers and more than 8000 journals in the CrossRef system and nearly 10 million DOIs for articles, book chapters and conference proceedings registered. The DOI reference links are being heavily used - in October 2003 there were over 5 million DOI resolutions. A resolution of a DOI is an actual user clicking on a reference in an online journal - so CrossRef is having an impact. CrossRef is on target for adding over 3.2 million DOIs in 2003. One of the reasons this is so high is that publishers are digitizing their backfiles at a great rate and many publishers are going back to volume 1, issue 1 for their journals. In fact, the oldest article in the CrossRef system is from 1849 from the *Astronomical Journal*. The fact that such an old article can be cited in a reference and linked to and accessed electronically is fantastic. Journal ownership changes have been much smoother with DOIs. When a journal changes hands, ownership of the affected DOIs is transferred to the new publisher who updates the DOIs with new URLs in the central DOI system. Anyone using the DOI will seamlessly go to new journal site.

Libraries and DOIs

CrossRef and DOIs can help libraries in a number of ways. Libraries should find, and demand, DOIs in licensed content and databases since it will mean easy full text links. Libraries can also add DOIs to any article-level information they hold locally - libraries can retrieve DOIs from publishers or directly from CrossRef at no cost. Libraries can send bibliographic data to CrossRef and get a DOI, or they can send a DOI and get standardized metadata. This is important because CrossRef and DOIs integrate with OpenURL Local Link Servers provided by vendors like Ex Libris (SFX), EBSCO (LinkSource) and Endeavor (LinkFinderPlus).

Solving the appropriate Copy Problem/OpenURL

The appropriate copy problem refers to the fact that a user may have access to full text from different services or locally hosted content. Right now, the DOI resolves to one URL registered by the publisher. The problem is that a user may click a link and go to the publisher's site and be asked to pay access to the full text. For example, a user could click a link at ACM and go to Elsevier and be asked to pay for the article while the full text article is available locally. This is because the link is originating at ACM and ACM doesn't know what copies of an Elsevier article are available locally. So, in general terms the appropriate copy problem is: when more than 1 copy exists, users frequently have the rights to access different copies. The solution to this involves OpenURL. The OpenURL framework is a system for enabling context sensitive, or localized linking. The OpenURL itself is a protocol for transporting metadata and identifiers (including DOIs) in a URL. NISO (National Information Standards Organization) is just finishing up the OpenURL 1.0 standard. There is a common misperception that DOIs and OpenURL compete - this is not the case. OpenURL is not an identifier and it does not provide a resolution service (two things that DOIs do provide). DOIs don't provide localized links based on an institution's holdings. Combining DOIs and OpenURL achieves the best of both worlds - standardized metadata, unique identification of articles and localized links. An OpenURL with a DOI in it is more accurate than an OpenURL with just bibliographic metadata in it. This is why CrossRef and DOIs are OpenURL-aware. When a user with a local link server clicks a DOI, the DOI will be redirected to the local link server, not the publisher's site. The local link server sends the DOI to CrossRef and gets authoritative, clean metadata from CrossRef to construct appropriate links for users.

LATEST DEVELOPMENTS

CrossRef started with journal articles and is adding conference proceedings and book content. In addition, while STM publishers started CrossRef, members now include large and small publishers from all areas of scholarly communications. Discussions have also started about adding theses and dissertations. In addition there is a whole area of gray literature to consider.

CrossRef is working on what is called "forward linking". Starting in 2004, when publishers register a DOI and the metadata for an article, they will also deposit the references for the article. This will enable CrossRef to connect an article to other articles that cite it and provide "cited by" links. It will be optional for publishers to provide their references for this service. The service will be used by publishers to add "cited by" links to their websites. Many publishers are already providing "cited by" links, but only within one system or platform (HighWire Press, IOP) - the CrossRef service will potentially extend this functionality across many journals.

CONCLUSION

In the online world collaboration and standards are necessary to meet end users' ever growing demands. CrossRef provides an organization for publishers to collaborate with one another on reference linking, but it also enables publishers to work with libraries. CrossRef worked closely with libraries and library vendors on the solution to the appropriate copy problem by integrating DOIs, CrossRef and OpenURL. We hope to continue this type of collaboration in the future.

REFERENCES

Jones, Steve. *The Internet goes to College*. Pew Internet and American Life Project. September 15, 2002. http://www.pewinternet.org/reports/pdfs/PIP_College_Report.pdf

WEB SITES REFERRED TO IN THE TEXT

BioMed Central. http://www.biomedcentral.com/

CrossRef. http://www.crossref.org/

DOI - Digital Object Identifier system. http://www.doi.org/index.html

DOI Directory. http://dx.doi.org/

IDF - International DOI Foundation. http://www.doi.org/

LinkFinder*Plus*. http://www.endinfosys.com/prods/linkfinderplus.htm

LinkSource. http://www.linkresolver.com/

NISO - National Information Standards Organization. http://www.niso.org/index.html

PLoS - Public Library of Science. http://www.publiclibraryofscience.org/

SFX. http://www.sfxit.com/