

CrossRef

The missing link

by Ed Pentz

Citing articles in references is one of the foundations of the scholarly communication system. With references, authors make explicit links between their research and other articles that may, on the surface, appear unrelated.

Eugene Garfield, the founder of ISI, tells us that through references, "authors should formally assert and verify their ideas are original and do not replicate discoveries already reported in the archive."¹ Links enable users to see the body of primary literature as an interconnected collection of articles. When we can move from a reference to the full text of a cited article in one or two clicks, we will truly realize the benefits of electronic publication.

With the advent of electronic journals, a lot of attention has been given to multimedia capabilities and features like online peer review, but a full system of reference links is an essential feature that has been missing to date. Peter Boyce of the American Astronomical Society (AAS), reporting in the pioneering online journal "Astrophysics Journal Letters," wrote in 1997, "Reader feedback continues to emphasize the importance of links by which it is possible to retrieve referenced articles . . ."²

Because reference linking is so important, publishers of scholarly journals have an economic imperative to provide reference links—

journals without links will be seen as less valuable or useful than those with links.

Many online journals have some links and have had them for a number of years. However, most of this linking has been within a very narrow, focused subject area, between large secondary database publishers and large primary publishers, or within proprietary journal systems.

For example, the astronomy literature is very well linked through the Astrophysics Data System (ADS)³ and HighWire Press⁴ has extensive reference links between HighWire journals and PubMed.⁵ What has been missing is direct links between primary publishers and links between secondary and smaller publishers. Reference linking has been held back by the need for bilateral linking agreements between individual publishers; drafting such agreements is a laborious and time-consuming legal process. To have abundant links, a publisher would have to sign agreements with hundreds of organizations, an unworkable proposition. It is especially difficult for smaller publishers without extensive staffs to participate in reference linking.

CrossRef and PILA

Aware of the importance of linking and of the inefficiency of signing bilateral linking agreements, publishers took the unusual step of cooperating to set up CrossRef, a collabor-

About the author

Ed Pentz is executive director of CrossRef, Publishers International Linking Association, in Burlington, Massachusetts, e-mail: epentz@crossref.org

Editors' introduction

As academic librarians, we are all too familiar with the challenges that researchers face in following ideas through a long chain of citations.

From article to shelf to microfilm to InterLibrary Loan and back again, the process of tracking down articles that form the pieces of an intellectual puzzle has often been a difficult one. One faculty member I know used to refer to the process as a "lifetime of bibliographic frustration."

In contrast, one of the joys of information on the Web is the ability to move quickly

through a series of related sites. With the advent of electronic journals, many of us expected to see journal citations transformed into instant links from full-text article to full-text article, regardless of publisher. This transformation has been slow to arrive, for a variety of technical, economic, and legal reasons.

Our column for this issue describes an important effort to make electronic linking a reality: Cross Ref.—*Ivy Anderson, Gail McMillan, and Ann Schaffner*

reference linking service. At the end of 1999, a group of leading scientific, technical, and medical (STM) publishers joined to form the nonprofit, independent organization, Publishers International Linking Association, Inc. (PILA), which operates CrossRef from a central location in Burlington, Massachusetts.

The PILA Board of Directors includes representatives from AAAS (Science), American Press (Harcourt), American Institute of Physics, Association of Computing Machinery, Blackwell Science, Elsevier Science, Institute of Electronics and Electrical Engineering, Kluwer, Nature, Oxford University Press, Springer Verlag, and John Wiley & Sons.

Even though CrossRef was incorporated in January 2000, the CrossRef system is already up and running. There are well over 60 publishers participating in CrossRef, accounting for nearly 3,000 journals with about 2.1 million article records. More than 60% of CrossRef members are nonprofit publishers, and, while STM was an initial focus, CrossRef now covers all areas of scholarly publishing.

CrossRef functions as a sort of digital switchboard. It holds no full-text content, but creates linkages through Digital Object Identifier (DOI) numbers, which are tagged to article metadata supplied by the participating publishers. A researcher clicking on a link will be connected to a page on the publisher's Web site showing a full bibliographical cita-

tion of the article, and, in most cases, the abstract. The format of the link is determined by publisher preference; for example, a CrossRef button or "Article" in html. The reader can then access the full-text article through the appropriate mechanism; subscribers will generally go straight to the text, while others will receive information on access via subscription, document delivery, or pay-per-view.

It is important to note that CrossRef acts "behind-the-scenes" and collects only a minimal amount of bibliographic metadata. Abstracts and full-text articles remain at publishers' sites, and access to the material is controlled by publishers' access control systems. This has been referred to as "distributed aggregation." Users who are subscribers to the cited journal will in most cases have their Internet Protocol (IP) address checked and be able to access the full-text content seamlessly. CrossRef is not a search system. End users do not access CrossRef directly; organizations access CrossRef to look up DOIs to create full-text links to scholarly journal articles.

Creating and linking to DOIs

For participating publishers, CrossRef offers three main services: the depositing of article metadata in the CrossRef database, the submission of the references in those articles for

About the editors

Ivy Anderson is coordinator for Digital Acquisitions at Harvard University, e-mail: ivy_anderson@harvard.edu; Gail McMillan is director of Digital Library and Archives at Virginia Polytechnic Institute and State University, e-mail: gailmac@vt.edu; Ann Schaffner has been an academic librarian for more than 20 years and is currently a full time MBA student at Simmons College, e-mail: ann.schaffner@simmons.edu

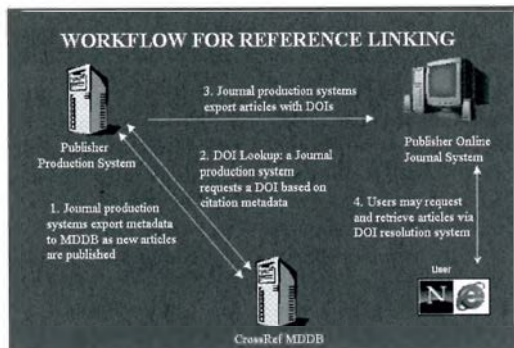


Figure 1: "Workflow for DOI Reference Linking" **crossref**

the purpose of obtaining their DOIs, and the creation of links using those DOIs.

The first step is for the publisher to obtain a DOI prefix from the International DOI Foundation (IDF) at <http://www.doi.org>. The cost of this service is covered by CrossRef membership. The publisher then submits minimal article metadata (journal title, article title, volume, issue, page, and first author) to the metadata database (MDDB) with the DOI and URL. Metadata is to be in an XML-based Document Type Definition format, the standards of which are provided on the CrossRef Web site (<http://www.crossref.org>). As part of the submission process, CrossRef registers the article's DOI and URL in the central DOI Directory, run by the IDF.

The publisher then submits the reference citations contained in each journal article to the Reference Resolver (RR), a front-end component of the MDDB. The RR allows the retrieval of DOIs, enabling the publisher to create links. The format and protocol for these submissions are also covered on the CrossRef Web site. The publisher uses the DOI to create a normal DOI link (see Figure 1). The DOI is sent to the DOI Directory and automatically resolved to the URL deposited by the publisher (see Figure 2).

An example of a DOI is 10.1006/jmbi.1995.2434—it is for an article from Academic Press's *Journal of Molecular Biology*, available on the IDEAL system. "10.1006" is Academic Press's prefix (each publisher has a unique prefix). After the "/" the publisher determines how to identify the article. In this case Academic Press

uses a four-letter code for the journal, the year of acceptance, and a sequential article number. This DOI as a link would appear as: <http://dx.doi.org/10.1006/jmbi.1995.2434>. Clicking on the link will take the user to the abstract page on the IDEAL system. Some publishers are using SICIs (Serial Item and Contribution Identifier)⁶ or PIIs (Publisher Item Identifiers)⁷ for their DOIs.

The DOI is a very powerful tool. Reference linking until now has depended largely on algorithmic links, which employ URLs. Since a URL is

not a true identifier, but a pointer to a location on a particular machine, one can reach a "file not found" error message if the file has been moved. A more serious problem is that this approach, like bilateral agreements, is not scalable; every publisher has to know and track changes in the linking format of every other publisher, which becomes an overwhelming task as linkage proliferates.

By taking the standards-based DOI approach, in which a given DOI is always associated with a specific article, CrossRef has removed the need for participants to archive linking schemes. If a publisher changes its URL, only the central DOI Directory needs to be updated and each DOI will automatically resolve to its new URL.

The International DOI Foundation ensures interoperability among different user communities. Through close cooperation with IDF, CrossRef has launched the first large-scale, practical DOI application to address the sophisticated demands of readers of scientific and scholarly journals. CrossRef has

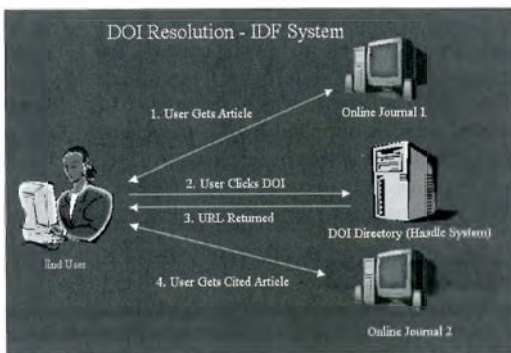


Figure 2: "DOI Resolution" **crossref**

also become the first official DOI Registration Agency, granting it the means to assign DOI prefixes to CrossRef members, and to register DOIs in the system.

As a collaborative venture, the success of CrossRef depends on the cooperation of its members. Publishers must be prepared to receive incoming links at the time of metadata submission. They are also expected to maintain the accuracy of their metadata, DOIs, and URLs, and to provide information on article access.

CrossRef membership is open to primary scholarly publishers. However, many other organizations can benefit from using CrossRef to look up DOIs to create links to full-text articles. To fill this need, CrossRef has created an affiliates category. Affiliates are non-members, such as secondary database producers, subscription agents, and abstracting and indexing services who can sign up to use the CrossRef system.

CrossRef costs the researcher nothing; its expenses are covered by charges to member publishers for depositing their metadata, retrieving DOIs, and annual membership fees. There are no charges for clicking on links. Affiliates pay an annual administrative fee and retrieval fees for looking up DOIs. Library affiliates can pay a flat fee of \$500 for unlimited access to DOI look up. Current fee schedules are posted on the CrossRef Web site. Fees are designed to cover costs based on use of the system (so small publishers pay lower fees than larger publishers do). CrossRef itself has no stake in publishers' decisions regarding their charges for content access.

Inevitably, problems unique to the digital realm have arisen. Of most concern to libraries is what is known as the "appropriate copy" issue. Since a user at an institution may have access to a given article through more than one source, he or she must be able to discover which is the "appropriate" copy.

For example, a library user should not pay for an article at the publisher's Web site if it is also available through a library subscription to Ovid, EBSCO Online, or in the library's print holdings. The question of how to provide "localized links" so that users can get to appropriate copies has been under discussion for several years.⁸

To move this process along, CrossRef co-sponsored the "Workshop on Localization in Reference Linking" with NISO, DLF, CNRI, and IDF.⁹ At this meeting a general architecture for localized linking was outlined and a practical prototype of this type of linking is now being planned. The prototype will involve CrossRef, IDF, DLF, publishers, libraries, and others working together. DOIs, metadata, and OpenURL are all important parts of the localized linking solution and they all work together. CrossRef is committed to working with libraries and others on solutions to these problems.

Another major issue is the crucial question of how digital content is to be archived. Here, too, CrossRef is seeking the answers we will all need in the years ahead. For example, CrossRef hopes to link to such archiving systems as JSTOR, which scans journal issues, in some cases going back to the 1800s. Assigning DOIs to these older articles means that they can be included in the linking network. When a user can click on a citation to an article from the 19th century and get to the full text online, scholarly communication will truly be transformed.

CrossRef provides the "missing link" in linking, making broad-based linking efficient and manageable for large and small publishers. CrossRef is available for other organizations to use and can benefit the entire scholarly communications process. Because CrossRef has taken the approach of using open standards, it will need to be interoperable with other linking systems. The DOIs and metadata that CrossRef use lay the groundwork for more sophisticated linking in the future.

Notes

1. Eugene Garfield, "The Concept of Citation Indexing" *Current Contents*, January 3, 1994, <http://www.isinet.com/isi/hot/essays/1.html>.
2. Peter Boyce, "Electronic Publishing: Experience is Telling us Something." *Serials Review* 23, no. 1 (1997): 1-10.
3. NASA Astrophysics Data System, <http://adswww.harvard.edu>.
4. HighWire Press, <http://www.highwire.org/>.

(continued on page 228)

CURRENT, AUTHORITATIVE STATISTICS ABOUT THE UNITED STATES.

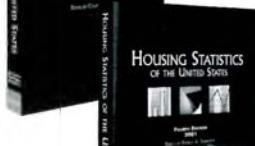
Whether you need data on foreign trade...



May 2001. Paperback. Approx. 450 pages.
ISBN: 0-89059-249-7. \$147

**Forthcoming
2nd Edition**

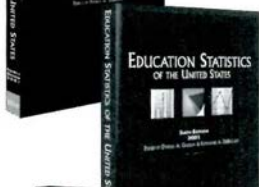
...or housing...



May 2001. Paperback. Approx. 450 pages.
ISBN: 0-89059-284-5. \$147

**Forthcoming
4th Edition**

...or education...



April 2001. Paperback. Approx. 450 pages.
ISBN: 0-89059-360-4. \$147

**Forthcoming
3rd Edition**

...or business and economics...



February 2001. Paperback. Approx. 500 pages.
ISBN: 0-89059-282-9. \$147

...or health and healthcare...



January 2001. Paperback. Approx. 450 pages.
ISBN: 0-89059-278-0. \$135

...or labor and employment...



November 2000. Paperback. 423 pages.
ISBN: 0-89059-281-0. \$147

**Visit us at
Booth 106
at the ACR
conference
in Denver!**

...Bernan's U.S. DataBook Series™
has the information you need!

- Hard-to-find state, county, and metropolitan area level data!
- Data series no longer available in print from any other source!
- Insightful analysis from expert editors and economists!
- Source notes and definitions!

BERNAN

ESSENTIAL GOVERNMENT INFORMATION

Call toll-free: (800) 865-3457

Fax toll-free: (800) 865-3450

Email: order@bernan.com

www.bernan.com

Retirements

Julita C. Awkard has retired, after 38 years of service, as university librarian at Florida Agricultural & Mechanical University.

Deaths

Perry D. Morrison, 81, retired science librarian at the University of Oregon (UO), died December 7. From 1949 to 1963, Morrison was head social science librarian at UO. He left there in 1963 to become head librarian and director of the library science program at Sacramento State College (now Sacramento State University). From 1965 to 1967 he was on the faculty of the School of Librarianship at the University of Washington in Seattle. He returned to UO as a faculty member in its new School of Librarianship, later serving as its dean from 1971 until 1973. With the suspension of the school, he became a science librarian until his retirement in 1982. Morrison edited the Pacific Northwest Library Association's quarterly from 1967 to 1972. He

also served as president of the Oregon Library Association from 1961 to 1962. Morrison was published in many library publications and was for many years a contributor to *Collier's Encyclopedia Year Book* about Oregon libraries, enjoyed reviewing professional library publications, and wrote articles about local history. ■

(*"Cross Ref" continued from page 209*)

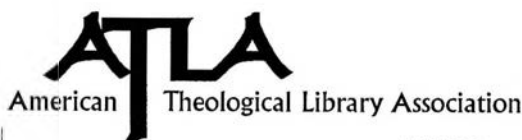
5. PubMed, <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=PubMed>.

6. ANSI/NISO Z39.56-1996 Serial Item and Contribution Identifier (SICI).

7. Lorrin Garson, "Publisher Item Identifier as a Means of Document Identification," <http://pubs.acs.org/journals/pubiden.html>.

8. Priscilla Caplan and William Arms, "Reference Linking for Journal Articles" *D-Lib Magazine* (July/August 1999), <http://www.dlib.org/dlib/july99/caplan/07caplan.html>.

9. NISO/DLF/CrossRef Workshop on Localization in Reference Linking July 24, 2000—CNRI, Reston, Virginia, Meeting Report, <http://www.niso.org/CNRI-mtg.html>. ■



WIN

Visit our booth **#634** at the ACRL Conference and register to win a one-year subscription to

The ATLA Religion Database on CD-ROM

The premier index to the scholarly literature of religion

or

ATLASerials

The American Theological Library Association's major digital serials journal project

For more information on the American Theological Library Association's ATLASerials project, contact Chuck Slagle at cslagle@atla.com

Toll-free (North America): (888) 665-ATLA • <http://www.atla.com>