Feature article

CrossRef and SFX: complementary linking services for libraries

Jenny Walker

The author

Jenny Walker is Director, Sales and Marketing, Ex Libris (USA) Inc., Watertown, Massachusetts, USA.

Keywords

Libraries, Library services, Links, Electronic data, Electronic journals, Standards

Abstract

With the increase in the use of electronic information services in libraries, and in particular with the dramatic increase in the use of electronic journals, there is an urgent need by libraries for solutions that link the disparate information resources in a meaningful way for the end user and that optimize the use of these resources. Such linking solutions are now available for libraries, supported and assisted by the emergence of new standards such as the OpenURL and the Digital Object Identifier (DOI). Linking solutions built around these standards include SFX and CrossRef. Demonstrates how these different solutions, and the underlying standards, interact to meet library needs.

Electronic access

The research register for this journal is available at
http://www.emeraldinsight.com/researchregisters

The current issue and full text archive of this journal is available at
http://www.emeraldinsight.com/0307-4803.htm

New Library World
Volume 103 · Number 1174 · 2002 · pp. 83–89
© MCB UP Limited · ISSN 0307-4803
DOI 10.1108/03074800210422296

Introduction

Priscilla Caplan's (2001) article "A lesson in linking" has been one of the most useful of many recent articles on linking, now a very "hot" topic for libraries – and also for publishers. Libraries wish to integrate – link – in a meaningful way the many information resources in which they invest significant financial resources. Publishers, too, favour such integration as they wish to enhance their services, attracting as much traffic as possible to their sites.

The most common form of linking among scholarly information resources is from a citation to the "item" which is referenced by the citation, typically the book or the article. However, many other possible linking services will help leverage the inherent value of the resources.

A brief history of linking

Linking in an electronic environment is not a new concept, although it is now becoming a reality for many involved in scholarly research. Vannevar Bush (1945), in his ground-breaking paper "As we may think", identified the importance of actionable links between conceptually related items. Bush described an approach he termed "associative indexing" and suggested that this process of selecting by association rather than through indexing might be mechanized. The Web has not yet shown itself flexible enough to support Bush's vision to link everything to everything under the control of the individual user, but breakthrough solutions are emerging providing control – at least at the level of the library.

In the mid-1990s, vendors of abstracting and indexing (A&I) databases introduced links from their systems to library OPACs so that users, having retrieved a citation, could check local library holdings for print copies of the journal containing the cited article. To this, vendors added links to e-journals as these became available. However, deployment was inconsistent and incomplete – A&I vendors offered full text linking in different ways and were dependent in most cases on agreements with publishers. Not all A&I vendors were able to sign agreements with all e-journal publishers and aggregators.

New Library World
Volume 103 · Number 1174 · 2002 · 83–89

An alternative approach to linking between resources from different vendors was to republish electronically a clearly defined set of journals and provide interlinking within this discrete set of data. Such an approach was taken by Ovid with their *BioMedical Collections on CD-ROM*, now available through Journals@Ovid. This has been very successful in certain niche markets.

Scalability, robustness and usability of such vendor-specific linking solutions as those mentioned can be difficult and so we see new "open" linking solutions emerge.

Basic building blocks

According to Caplan (2001) the basic building blocks of a linking system are identifiers, resolvers and reference databases.

In the reference-linking environment, significant effort has been put into the development of identifiers that can uniquely identify an object and also that can be understood by heterogeneous systems.

Much in the electronic world has developed from the paper paradigm, and as far as identification or naming schemes are concerned, these tend, for journals, to centre on journal, volume, issue, and page number, e.g. serial item contribution identifier (SICI). Had we been starting out now with such naming schemes we would likely have started with the work itself and cascaded down to digital and physical manifestations. It is important to identify each aspect – for citation purposes an author of a research article would want all instances of his/her work to be counted irrespective of whether it is printed in a paper journal, or published in an electronic journal on the Web – i.e. irrespective of the manifestation. However, a researcher may want a particular manifestation of that work and require separate identification.

No single identifier is likely to meet all needs. What is required is a network of identifiers, with supporting metadata about the entity. The Digital Object Identifier (DOI) is one such identifier and is that most favoured by publishers.

Metadata associated with an identifier is typically stored in a reference database. However, not all identifiers require such a database. A SICI identifier, for example, does not require associated metadata, as the SICI itself can be broken down into component parts each of which are meaningful and represent metadata. For "dumb" identifiers, such as the DOI, the reference database provides the mechanism to perform a look-up – to retrieve the identifier for any given citation or to retrieve the metadata for any given identifier. These look-up mechanisms become very important in solving the "appropriate copy" problem (Caplan and Arms, 1999) and in providing a range of extended linking services that go beyond the links to full text for a given citation (Van de Sompel and Hochstenbach, 1999).

While a reference database associates identifiers with metadata, a resolver, or resolution system, associates identifiers with location – specifying one or more locations for the content associated with an identifier. For an identifier to be usable in reference linking, there must be at least one resolution system for it. The International DOI Foundation (IDF) runs a resolver for DOIs. The DOI resolver is the handle system.

CrossRef linking

CrossRef is an initiative of the Publishers International Linking Association (PILA), and was formed in January 2000. CrossRef, started by 12 major publishers to enable linking between their resources, can now boast 92 member publishers – a remarkable achievement for the publishing community. CrossRef is one of the registration agencies for the IDF, which means that CrossRef is authorized to register DOIs on behalf of its members.

Journal publishers participating in CrossRef are assigned a unique DOI publisher prefix so that a DOI is unique across publishers. The publisher assigns a DOI to each article they publish. This DOI, together with basic metadata, is deposited in the CrossRef reference database, and the DOI resolver is notified of the location of the article – typically the publisher's own Web service or one operated on their behalf, e.g. by Highwire Press or Ingenta. In addition, as part of the publication process for new articles, the metadata for each reference in the article is used to query the CrossRef database. Where a match is found, the DOI is retrieved and inserted as a CrossRef link in the article reference. Thus, the beginnings of a very effective linking system are evident, which in

time will grow to cover a wider area of scholarship than is the case today, where some of the publishers are still at the early stages of implementing CrossRef links.

Users finding a CrossRef link will be able to link directly to the full text of the article referenced by that CrossRef link. The location of the article to which the user will be linked is that determined by the publisher of the article referenced. This would not be a problem if there were only ever one possible service – the publisher's own – where such full text could be found. However, the reality is that the electronic full text of the article could be found on some other service such as an aggregator service – or the institution could host the text locally. Furthermore, the article may be available at a particular institution only in print form and could be found on the library shelves. So how when a user clicks on a CrossRef link does the system take them to the "appropriate copy" (Caplan and Arms, 1999)?

Linking localization

Value judgements on "appropriateness" are the preserve of the librarian and are determined at the institution level; i.e. in the context of the collection (digital and print) of the institution. To support localization, institutions require a resolver that has knowledge about the local institution's collection. Centralised resolvers, such as the DOI Resolver, do not have any knowledge of an institution's collection. However, local resolvers have emerged that can be configured to reflect the library's collection. The first of these local resolvers – link servers – to be launched was the SFX link server from Ex Libris (http://www.sfxit.com). SFX was first developed by Herbert Van de Sompel and his team at Ghent University in Belgium and was purchased in February 2000 by Ex Libris. Other vendors have since announced similar link server solutions.

A link server, being under the control of the local institution, can exert complete control over the types of link services that are offered to users at any time and to where these are resolved. In this new world of linking, the link source, such as an article reference, includes not a static or hard coded link, but rather a "hook" which is inserted dynamically at the time the user views the link source, such as an article reference, and which describes the link source rather than the link target. A standard has already been defined for these linking "hooks" – the OpenURL.

The OpenURL

The OpenURL framework (Van de Sompel and Beit-Arie, 2001), is derived from the SFX research (Van de Sompel and Hochstenbach, 1999a, b, c), and is the architecture for the localized and context-sensitive resolution of metadata and identifiers of referenced scholarly works into services. It defines a model for open linking by which the provision of linking services for a referenced work is disconnected from the reference to the work, as presented to users in electronic information resources.

The OpenURL standard allows for descriptive metadata elements and/or identifiers to be transported from the link source to the link server. The OpenURL is implemented by the information service provider (ISP) in their service and can be applied by them at various levels within their resource. The OpenURL could, for example, in an abstracting and indexing database, be offered to the user for each record shown in a brief results list and/or for each full record shown. When a user clicks on a button in the link source, an OpenURL is created and sent to the link server. An SFX button represents the OpenURL sent to an SFX server. The OpenURL has two key components:

(1) The base-URL that determines the address of the link server to which the OpenURL should be sent. Sites must provide this information to the ISP for each link source. The notification can be automated.
(2) The query or "content". This portion of the OpenURL can contain metadata elements that describe the object or link source and/or identifiers that reference the link source.

The OpenURL, already a *de facto* standard for linking, is on a fast track path towards acceptance as a NISO standard (see http://library.caltech.edu/openurl). Pending acceptance by NISO, many ISPs have already implemented the OpenURL according to the published draft standard, and many others have committed to do so

(see http://www.sfxit.com/sources.html). The OpenURL is easy to implement, and it provides a cost-effective means of providing libraries with appropriate ways of linking across information resources irrespective of provider. ISPs, including CrossRef member publishers, can offer the OpenURL open linking capability alongside their default, resource-specific links and CrossRef links. For example, the Institute of Physics, while offering CrossRef links, also offers SFX links to those sites who have their own SFX link server (see Figures 1 and 2). In many cases there is no CrossRef link to take the user to the full text, and even where a CrossRef link is shown the user may not have access to the article. However, an SFX button will appear against each and every reference and may lead the user not only to the full text – perhaps from an aggregator database – but will provide library-configured options to lead the user to the print copy if available, or to facilitate an immediate document delivery request that does not require the user to log into another system and re-key the request details. Furthermore additional relevant services could be available.

Link Servers such as SFX can offer extended service links – links that go beyond linking to the full text. Many of the more creative services depend on comprehensive metadata being provided to the link server via the OpenURL. If, for example, the metadata on the OpenURL consists only of an identifier the link server would be severely handicapped in terms of the service links it could offer. However, if using that identifier the link server could retrieve the metadata from a reference database, a richer range of services could be offered to the user.

Where CrossRef and SFX intersect

Early in 2001 an important prototype was started to demonstrate the way in which the OpenURL framework for open linking could be integrated with the DOI/CrossRef linking solution. This prototype set out to address the "appropriate copy" problem whereby the user's context is not taken into consideration in the delivery of CrossRef links, leading to immense user frustration. Those participating in the prototype included important groups such as CrossRef, IDF, US Digital Library Federation (DLF), the Corporation for National Research Initiatives (CNRI) and Ex Libris (Van de Sompel and Beit-Arie, 2001; Beit-Arie *et al.*, 2001).

The solution demonstrated during the prototype, with the support of many publishers, was based on implementing the

Figure 1 The Institute of Physics (IOP) Electronic Journals Service provides CrossRef links in article references. However, as can be seen above, there are few CrossRef links available to the user, while SFX buttons are available for each reference. Over time the number of CrossRef links in article references will increase

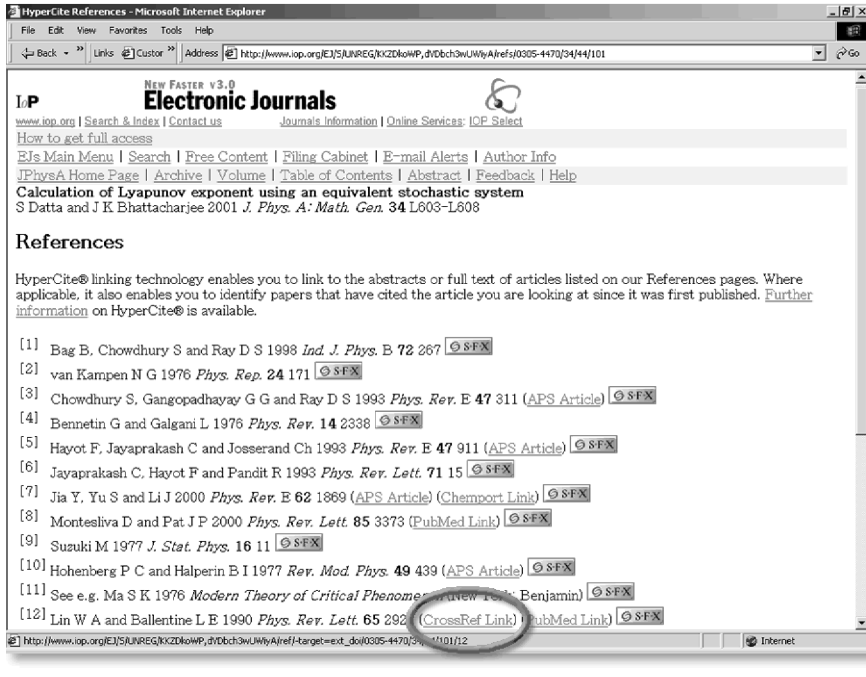
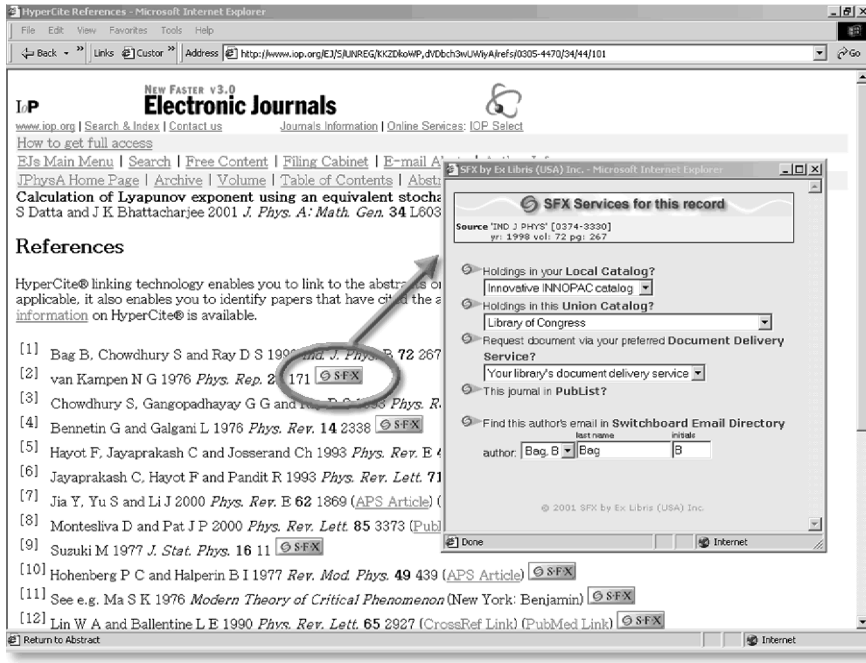


Figure 2 Clicking on the SFX button, even for a reference with no other IOP-provided links, will provide a number of service options for the user. In this example the options include the ability to check the local catalogue for a print version of the article, or to request the article via the library's preferred document delivery supplier, or via the institution's ILL system



OpenURL standard for the DOI/CrossRef environment. The simple concept on which the prototype was based was that the DOI proxy server, through which all DOI links pass, would detect whether or not a user clicking on a DOI (typically presented to the user as a "CrossRef" link) was a user from an institution with an OpenURL-compliant link server (Figure 3). If so, then the link would not be resolved by the DOI handle resolver but rather by the institution's own link server. As the link server has knowledge about the library's collection, only relevant links would be presented to the user and all such links would resolve to the "appropriate" copy or service.

Underlying this simple concept is a reference database look-up. The link server receives an OpenURL (from the DOI service) that contains only a DOI identifier. To deliver services the link server needs metadata and so retrieves this from the CrossRef database (see Figure 4).

Since all DOI links go through the DOI proxy server, and it is this central spot – the DOI server – that becomes OpenURL enabled, individual CrossRef publishers do not need to change their behaviour, yet library users will benefit from localized links.

The services demonstrated during the pilot project will now be extended to all CrossRef Library affiliate members.

CrossRef Library Affiliate Program

Users from CrossRef affiliate member libraries who also have an SFX or other OpenURL-compliant link server, will, when clicking on a CrossRef link, be directed via their local link server to the "appropriate copy" of the article as well as being offered a range of other extended "appropriate" services.

A further advantage of CrossRef Library affiliate membership is that member libraries may query the CrossRef reference database with citation metadata to retrieve a DOI. This capability allows for improved linking capabilities to some publisher sites. For example certain publishers require a DOI to link to the full text of the article. Often the DOI is not known to the user or to the SFX server, in which case the SFX server may use the metadata it has describing the object to query the CrossRef reference data to retrieve the DOI so that an effective link can be made. Some publishers that require DOIs for linking

Figure 3 Users clicking on a CrossRef (DOI) link embedded in an article reference will be directed, via the DOI resolver, to the full text of the article at the publisher's site, in this case Blackwell's Synergy service. The user may be denied access to the full text – or given an option to purchase the article – if their institution does not have a subscription to the journal via the Synergy service

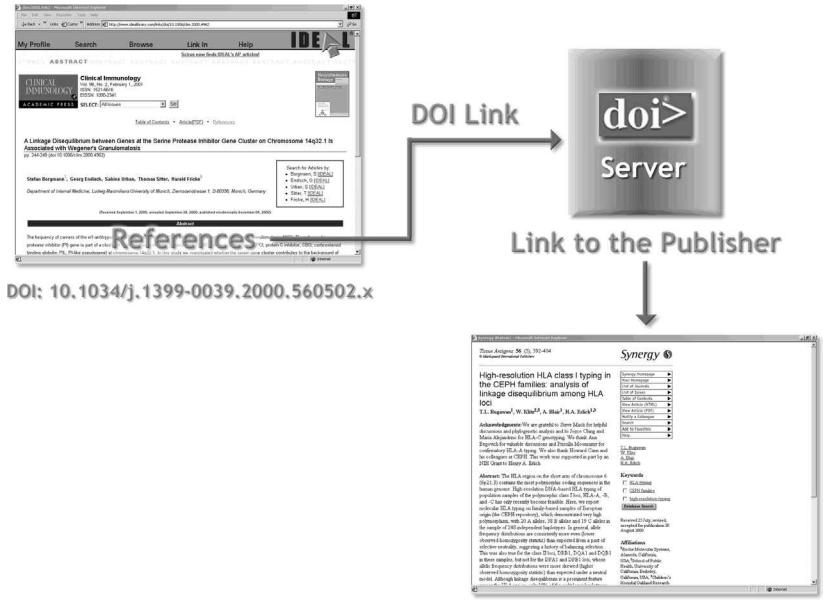
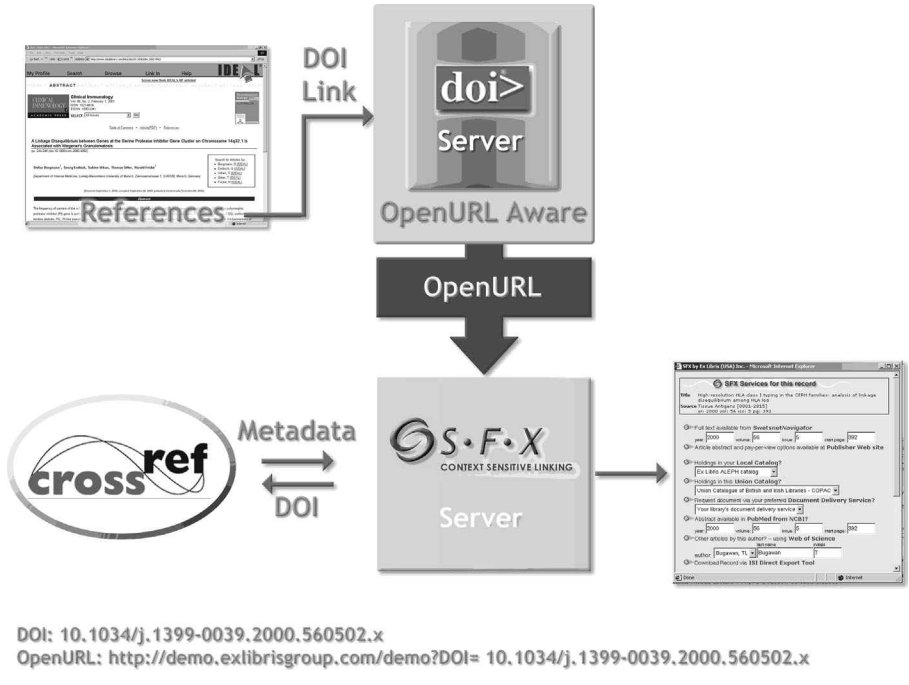


Figure 4 If the user's institution has an SFX link server, then the DOI server will recognize this and redirect the DOI – as an OpenURL – to the institution's SFX server. The SFX server will use the CrossRef database to retrieve the article metadata and can use this to determine the appropriate SFX services. In this example the "appropriate" full text of the article is available via SwetsNetNavigator. Extended service options are also offered to the user via the SFX service menu



include Elsevier Science, IEEE, APA and Wiley.

Further details on the CrossRef Library Affiliate Program can be found at http://www.crossref.org/library.htm . General information on CrossRef can be found at http://www.crossref.org, and in the excellent article, "CrossRef turns one", by Amy Brand of CrossRef, published in *D-Lib* in May 2001 (Brand, 2001).

New Library World
Volume 103 · Number 1174 · 2002 · 83–89

Complementary not competitive

The OpenURL and CrossRef linking frameworks can provide complementary services offering valuable solutions for libraries. The work that has been done during 2001 demonstrates the importance placed by all the stakeholders on finding ways to integrate the many and varied library resources in a meaningful way for library users.

References

Beit-Arie, O. et al. (2001), "Linking to the appropriate copy – report of a DOI-based prototype", D-Lib *Magazine, Vol. 7 No. 3, available at: www.dlib.org/* dlib/september01/caplan/09caplan.html

Brand, A. (2001), "CrossRef turns one", D-Lib Magazine, Vol. 7 No. 5, available at: www.dlib.org/dlib/may01/ brand/05brand.html

Bush, V. (1945), "As we may think", Atlantic Monthly, available at: www.isg.sfu.ca/~duchier/misc/vbush/ vbush-all.shtml

Caplan, P. (2001), "A lesson in linking", Library Journal, supplement, Fall, pp. 16-18.

Caplan, P. and Arms, W.Y. (1999), "Reference linking for journal articles", D-Lib Magazine, Vol. 5 No. 7/8, available at: www.dlib.org/dlib/july99/caplan/ 07caplan.html

Van de Sompel, H. and Beit-Arie, O. (2001), "Open linking in the scholarly information environment using the OpenURL framework", D-Lib Magazine, Vol. 7 No. 3, available at: www.dlib.org/dlib/march01/ vandesompel/03vandesompel.html

Van de Sompel, H. and Hochstenbach, P. (1999a), "Reference linking in a hybrid library environment. Part 1: frameworks for linking", D-Lib Magazine, Vol. 5 No. 4, available at: www.dlib.org/dlib/ april99/van_de_sompel/04van_de_sompel-pt1.html

Van de Sompel, H. and Hochstenbach, P. (1999b), "Part 2:SFX, a generic linking solution", D-Lib Magazine, Vol. 5 No. 4, available at: www.dlib.org/dlib/ april99/van_de_sompel/04van_de_sompel-pt2.html

Van de Sompel, H. and Hochstenbach, P. (1999c), "Part 3: generalizing the SFX solution in the 'SFX@Ghent and SFX@LANL' experiment", D-Lib Magazine, Vol. 5 No. 10, available at: www.cs.cornell.edu/ people/herbertv/paper/%20http://www.dlib.org/ dlib/october99/van_de_sompel/ 10van_de_sompel.html