

Introduction

Scope of This Report

Imagine a world where library users never reach a dead end, never fail to find the electronic resources they have the need—and the right—to access. This is the ultimate potential of the OpenURL: to link, seamlessly, among a multitude of information providers—proprietary and open access alike. In a world where libraries must acquire, manage, and provide access to a host of search tools and information sources from a variegated group of content providers, the promise of the OpenURL is tantalizing indeed.

Although such seamless integration of resources, via the OpenURL or other linking technologies, is tantalizing for libraries and information centers, it is simply *expected* by users. Herbert Van de Sompel and Patrick Hochstenbach, in the first of their 1999 three-part article series about reference linking and the OpenURL in *D-Lib Magazine* note: “When using a library solution, the expectations of a net-traveler are inspired by his hyperlinked Web-experiences. To such a user, it is not comprehensible that secondary sources, catalogues and primary sources, that are logically related, are not functionally linked.”¹ It is, fundamentally, the librarian’s job to facilitate the functionality that links these logically related resources.

A user should not have to leave one system and essentially re-create his or her search in another to find the full-text gold at the end of the rainbow. Users should not have to understand the complicated licensing relationships that allow them to view one copy of the full text and not another. As Van de Sompel and Hochstenbach comment, it is not only incomprehensible to users that all these sources are not linked, but the very power of the

electronic environment means that they *should not have to* comprehend it.

Of course, all this is much easier said than done. It is not only the technology that causes roadblocks but also the politics of all the information providers involved. Priscilla Caplan, in a 2001 *Library Journal* article, eloquently stated, “The idea of reference linking is so conceptually simple it is almost intuitive, but robust, scalable, open systems of reference linking are enormously complex,

“Reference Linking in a Hybrid Library Environment, Part 1: Frameworks for Linking,” by Herbert Van de Sompel and Patrick Hochstenbach
www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html

“A Lesson in Linking,” by Priscilla Caplan
www.libraryjournal.com/article/CA177643.html

and the standards and technical infrastructure to support them are only now being developed.”²

Four years later, the standards and technical infrastructure about which Caplan wrote are well developed, including the OpenURL and its many implementations, and much has been accomplished in a relatively short period of time. In fact, so much has been accomplished that a major nonlibrary resource, Google Scholar, has embraced and made possible OpenURL linking within its search interface.

When discussing the electronic linking of scholarly materials, it is important to remember that much has occurred in a very short time. Before delving into the

OpenURL's development and implementation, it is wise to take a short trip down memory lane and look at linking before the OpenURL. Chapter II, "On the Road to the OpenURL," focuses on basic reference linking in the pre-OpenURL world and provides a few basic foundation concepts and definitions.

With chapter III ("Development of Context-Sensitive Linking"), enter stage left the OpenURL. In 2000, the grandfather of the link-resolver products, Ex Libris's SFX, was in beta testing at Los Alamos National Laboratory, Harvard University Library, California Institute of Technology, Vanderbilt University, and Indiana University Purdue University Indianapolis. As of June 2005, SFX had 847 library customers, and many other link-resolver products have flooded the market. What were the steps along the way? What is the technology behind the OpenURL and what is its relationship to CrossRef and the DOI? What is the difference between OpenURL version 0.1 and the new standard, version 1.0? This chapter examines the development of context-sensitive linking

Google Scholar
<http://scholar.google.com>

Ex Libris SFX OpenURL Overview
www.exlibrisgroup.com/sfx_openurl.htm

and the major players and technological innovations that make such linking possible.

Although SFX was the first commercially available link resolver, by no means is SFX the only game in town. There are many link-resolver products that can help libraries exploit the OpenURL to greater and lesser degrees. Chapter IV, "Link-Resolver Products," describes and explains the options for link resolvers: commercial and homegrown/open source.

On the other end of the spectrum, there are libraries that have not implemented a link resolver for whatever reason—perhaps due to lack of funding or staffing, or other circumstances. Content providers continue to offer alternatives for linking without a stand-alone link resolver as well as linking alternatives to use in tandem with a stand-alone link resolver. Chapter V, "Linking without a Stand-Alone Link Resolver," examines these vendor-supplied solutions for context-sensitive linking and how vendors are increasingly providing local control for linking configurations.

Exploiting the OpenURL, particularly v. 1.0, means linking to more than just the full text of scholarly journals. Chapter VI, "Innovative Uses of the OpenURL," canvasses how many librarians and other linking gurus are using the OpenURL in innovative ways, including linking to a variety

of extended services as well as exploiting the extensibility of v. 1.0 in developments such as ContextObject in Span (COinS).

Many other issues exist for libraries and context-sensitive linking, including Google Scholar and other freely available scholarly search engines, linking for open-access materials and digital objects, and linking initiatives at OCLC. Chapter VII, "Other Linking Issues," will explore a variety of miscellaneous ingredients that make the current linking milieu so exciting and challenging for libraries and content providers.

This report is intended to serve as a general survey of the OpenURL and context-sensitive linking. At the end of the report are resources (Chapter VIII, "Sources and Resources"), which delve more deeply into the technologies behind the OpenURL, such as detailed specifications for the OpenURL v. 1.0, as well as resources that discuss in more detail some of the topics only briefly considered here.

The Appropriate Copy Problem

If only one copy existed of any given resource, then the appropriate copy problem would not exist. If only one copy, electronic or otherwise, existed of a scholarly article, then that copy would be the de facto appropriate one. Things are not quite this simple, however. In addition to the multiple electronic availability of the same scholarly article from a variety of sources, such as direct from the publisher or from an aggregator, there is also the issue of versioning. In this report, the phrase "appropriate copy" is used to denote the copy of a resource to which a user has rightful access, usually by virtue of his or her institutional affiliations but perhaps also by virtue of an information object being *open access*.

For example, a user may have rightful access to copy A from aggregator A, but not to copy B direct from publisher B. Therefore, any links that the user follows, in an attempt to get to the full text, need to lead him or her along path A to copy A. This is one problem the OpenURL, via link resolvers, attempts to solve. For librarians, the worst-case scenario is the user—in an attempt to find the full text—being led astray from his or her appropriate copy path and subsequently paying (via pay-per-view or otherwise) for resources to which he or she has rightful, free access. This is not uncommon and is especially distressing in the case of financially challenged students and other patrons.

Versioning, on the other hand, is the availability (of multiple copies) of different versions of a resource. In 2005, Sally Morris, chief executive, Association of Learned and Professional Society Publishers (ALPSP), gave a partial list of version possibilities for one article:

1. Privately circulated early draft (could be >1 iteration);

2. Version presented at public event (again, could be >1);
3. Presubmission version(s);
4. Version as submitted to journal x (may differ when resubmitted to journal y);
5. Version amended after peer review (may go through >1 round of amendment);
6. Version as accepted by journal x;
7. Accepted version, with substantive editing by journal editor and/or publisher (again, may be multiple iterations);
8. Accepted version, with substantive editing and copy-editing—ready for publication;
9. Publication version (as above, formatted and paginated)—proof;
10. Publication version, corrected and passed for publication;
11. Published version, not on publisher’s site (e.g., PDF), thus potentially lacking some functionality;
12. Published version (on publisher’s site, with full functionality); and
13. Post-publication version with errata/addenda (may be on publisher’s site, with functionality), or elsewhere without it.³

Many of these types of versions are open access and prove to be a different kind of challenge for linking issues; these issues are explored in greater depth in chapter VII (“Other Linking Issues”).

The issues of appropriate copy and versioning are inextricably intertwined, and the vocabulary used for both is often one and the same. Stevan Harnad, a preeminent open access advocate, has stated, “For a published article, the appropriate copy is always the copy of record, at the publisher’s Web site.”⁴ In the sense of which version is appropriate for scientific record, Harnad is correct. But in terms of which copy of the article is appropriate for a user to access, the one at the publisher’s site may or may not be the appropriate one—because the user could have access to an equally valid copy from an aggregator.

Before descending too deeply into semantics and vocabulary quarrels, however, I need to address what is at issue—the difference between access and copy of record. For one resource, there literally may be many electronic copies and versions, some or all of which may be useful and appropriate for a particular user, depending on his or her immediate needs.

In addition to all the electronic copies discussed previously, the appropriate copy—or the copy to which the user has the rights to access—may be the one on the shelf at the local library; the one the user can receive for free or for a small fee via Interlibrary Loan (ILL); or the one a user can receive through his or her institution’s preferred document-delivery service. Again, this is an issue of access, *not* of copy of record, and link resolvers can point users to the local library catalog or local ILL and document-delivery services.

Without the multiple availability of any given resource, the ultimate need for the OpenURL would not exist, but there is no monopolistic master database. The publishing world, particularly in electronic publishing, is messy and complicated. Thus, the OpenURL fills a very real need: to point users, from wherever they begin their search, to the copy of any given resource to which they have the rights to access.

Notes

1. Herbert Van de Sompel and Patrick Hochstenback, “Reference Linking in a Hybrid Library Environment, Part 1: Frame works for Linking,” *D-Lib Magazine* 5, no. 4 (April 1999), www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html (accessed November 28, 2005).
2. Priscilla Caplan, “A Lesson in Linking,” *Library Journal* 126, no. 17 (November 1, 2005; fall 2001 *netConnect*): 16, www.libraryjournal.com/article/CA177643.html (accessed November 28, 2005).
3. Sally Morris, e-mail correspondence with the author, February 5, 2005.
4. Stevan Harnad, e-mail correspondence with author, February 1, 2005.