Chapter 1

Introduction

Abstract

This chapter of "Rethinking Library Linking" introduces the concepts and purposes of link resolver software and the OpenURL standard and how current user behavior and new tools worked in tandem to create change in what is required for an effective link resolver.

Scope of This Report

The January/February 2006 issue of *Library Technology Reports*[1] introduced the OpenURL standard, its history, and its purpose for addressing the "complexity inherent in having multiple online copies" of an article or other item, often in multiple sources.[2] An OpenURL link resolver is a software product that takes advantage of this standard to link a citation in one product to the item's full text, even if that full text exists within a different product. This report builds on its predecessor by outlining issues common to OpenURL resolver products and suggests ways that libraries can address them. This report is not an introduction to link resolver products and assumes basic knowledge about library databases and the online research process.

It's important to note that the authors' perspective is that of librarians passionate about enhancing the user experience by improving the tools that our libraries purchase, license, or build, not that of experts on link resolver software or on the OpenURL standard. The principles guiding this report include these:

- The resolver's main purpose is to "shorten the path" between citation and item.[3]
- The relationship between the library and the open Web, especially Google, must be complementary, not competitive.
- OpenURL and related or successive linking initiatives must be widely adopted inside and outside libraries to facilitate the best user access to scholarly content.
- OpenURL and other linking technologies must be efficient, effective, and transparent to the user.

This report provides practicing librarians with real-world examples and strategies for improving resolver usability and functionality in their own institutions. To prepare this report, the authors tested and evaluated link resolver installations at their own libraries. The Claremont Colleges Library subscribes to Serials Solutions' 360 Link, and EKU is a long-time customer of SFX, an Ex Libris product.

Why OpenURL?

OpenURL was devised to solve the "appropriate copy problem." As online content proliferated, it became possible for libraries to obtain the same content from multiple locales: directly from publishers and subscription agents; indirectly through licensing citation databases that contain full text; and, increasingly, from free online sources. Before the advent of OpenURL, the only way to know whether a journal was held by the library was to search multiple resources. Libraries often maintained direct links to electronic journal websites, either in the library catalog or in a simple HTML list. Potentially relevant citations

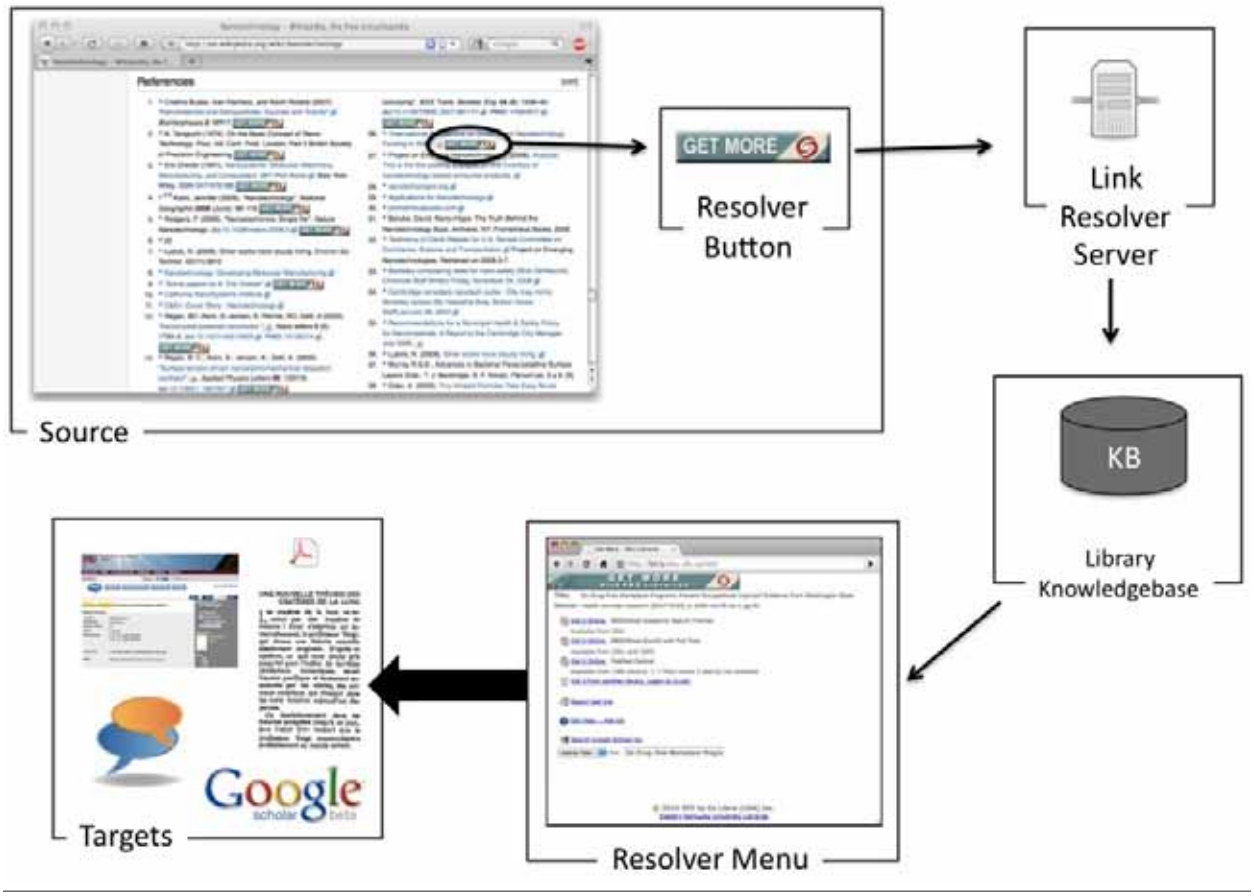

Figure 1
How OpenURL works.* PDF icon by Mark James. Used under a Creative Commons Attribution 2.5 Generic License.

were found in print and electronic indexes. Libraries have many indexes, referred to here as "citation databases," some of which may contain the full text of the items indexed therein. Full text items contained in a citation database are referred to in this report as "native full text." An OpenURL link resolver accepts links from library citation databases (sources) and returns to the user a menu of choices (targets) that may include links to full text, the library catalog, and other related services (figure 1). Key to understanding OpenURL is the concept of "context-sensitive" linking: links to the same item will be different for users of different libraries, and are dependent on the library's collections.

Basic Terms

These are some basic terms used in the discussion of OpenURL:

- **Aggregated database**—a citation database, often covering a wide or general subject area, that contains full text of some titles. The full text contained in such a database is negotiated by the database company (the aggregator) and is completely out of library control.
- **Base URL**—the Web address of a link resolver server for an institution. The base URL for a resolver must be known for library staff to OpenURL enable source databases.
- **Citation databases**—any online, searchable resource containing metadata for articles, books, book chapters, dissertations, reports, proceedings, and other items relevant to a user's topic. Citation databases are generally licensed by libraries for a fee.
- **Knowledge base**—the database describing the titles, availability dates, and URLs for all a library's holdings. A knowledge base is generally maintained by

* Wikipedia contributors, "Nanotechnology," Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/w/index.php?title=Nanotechnology&oldid=374711821 (accessed July 26, 2010).

the link resolver software vendor but is also customized by library staff to reflect variations in local holdings. For example, online access to some titles can vary by library, according to when the library first subscribed to the title or whether backfiles were purchased. Library staff typically add and maintain holdings data for journals and e-books, both individual items and packages, but aggregated database holdings are updated only by the link resolver vendor. Content creators supply link resolver vendors with metadata files, and link resolver vendors add these holdings to the knowledge base that drives holdings lists for all its customers.

- **Journal package**—a group of online journal titles purchased from a single publisher. Libraries may purchase multiple packages from a publisher. Packages often contain the most current content, necessitating the purchase of older holdings separately.
- **Link resolver**—software that interprets source OpenURLs, checks holdings in the local knowledge base, and creates links to targets and services. These links are presented in a Web browser window, which is generally called a resolver menu or the resolver results.
- **Native full text**—the complete text of articles or other items available in a source database. Native full text, in other words, is accessible in a citation database without the aid of an OpenURL link resolver.
- **OpenURL**—NISO standard Z39.88, by which Web links (URLs) are created containing bibliographic metadata, facilitating direct linking to articles, journals, books, chapters, dissertations, and more.
- **Source**—a citation database where an image or link to an OpenURL link resolver appears. There are many fewer sources than targets. Source databases are configured by libraries (e.g., Academic Search Premier) or by users (e.g., Google Scholar) and must comply with the OpenURL standard. Some citation databases are not OpenURL-compliant and therefore do not contain links to a library's link resolver.
- **Targets**—items linked from the resolver menu: native full text from a different source; publisher or electronic journal collection websites; the library catalog; Ask-a-Librarian; Google; and so on.

The OpenURL Process

Here is an outline of how the OpenURL process works (see figure 1):

- The user searches a source database and chooses a citation of interest.
- The user clicks a link or button embedded in that citation.
- An OpenURL is sent from the source to the library's link resolver.
- The OpenURL is interpreted by the link resolver.
- The link resolver checks the library's knowledge base.
- The link resolver determines if the data in the OpenURL meets the target's minimum requirements for creating an item-level link.
- If minimum requirements are met, a link directly to the item is presented to the user in menu form, along with related services. If the minimum requirements are not met, the resolver presents the next best link, sometimes to the issue's table of contents, the journal homepage, or (least preferably) to a database or publisher search page. Some resolver software presents multiple link levels as a safeguard against malformed or mistranslated article-level links.

The Appropriate Copy: Is It Still a Problem?

OpenURL link resolvers are still the best tool for the job of serving as middleman between diverse database resources and myriad full text locations that comprise library collections. However, as preprints, institutional repositories, and article-level open access grow, the capacity of knowledge bases to encompass the universe of potential appropriate copies is exceeded.

The "appropriate copy problem" is made more complex today by the open Web. Link resolvers cannot possibly track item availability across the entire open Web, though there are other linking initiatives that may help with this issue (see chapter 4). User and librarian opinions of link resolvers are compromised by this apparent gap. Related to the "appropriate copy" problem is the idea of "best copy." Many citation databases and publishers offer articles and other items in HTML as well as in PDF. This can be problematic when important information, such as figures, illustrations, and tables, is not available to users. It is important to take this into consideration when assigning rankings to targets that will govern the order in which they are presented to users.

Getting beyond "Appropriate Copy": Understanding Why OpenURL Resolving Fails

Link resolver users encounter two distinct categories of error, one obvious and one more hidden. A resolver returns a "false positive" error when it provides a link to an item that is not available in the library's subscriptions. These are the errors that are most often reported, since they reveal

themselves when a target link fails. The more hidden error, a "false negative," occurs when a resolver fails to link to an item that is in fact available. Because they are much less apparent to the user, false negatives can be more damaging to the user experience; if users subsequently find that a copy *is* available from the publisher or is openly available on the Web after not finding them with the help of their library's tools, users will lose faith in the efficacy of the resolver and, by extension, in their library.

These and other resolver errors can be traced to three main causes: source URL errors, target URL translation errors, and knowledge base inaccuracies. See chapter 3 for a full examination of each.

Tapping into the Power of Google Scholar

Resolver knowledge bases reflect title-level holdings for journals and books but cannot necessarily indicate whether individual articles are available. Because such is the case, we must at least provide users with an easy path to check the Web for item-level access in order to expand the universe of full text that is available to them via the resolver. Such content includes pre- and postprints in institutional repositories or individual articles made available via open access or as samples on publisher or author websites. At present, the best option for this appears to be Google Scholar. Operationally, the link to Google Scholar should be front and center whenever an OpenURL request does not provide a working knowledge-base-driven link to item-level full text. This is particularly important for book chapters and books, and Google Books results now appear in Google Scholar searches. Chapter-level requests sent to Google Scholar will frequently provide full text previews, with the entire chapter text being available in many cases. At the very least, these previews allow users to determine whether the item will meet their needs and allow them to request a print copy.

Google Scholar's deep indexing approach also frequently provides the most efficient means of access to publisher-hosted and open access content. Whenever a library's link resolver provides title-level rather than item-level access to this content, it will prove easier to access the item through Scholar, as long as it is contained in Scholar's index. Link resolvers need to take advantage of this more direct form of access to this growing component of the literature.

Discovery Tools: Shedding More Light on Link Resolver Failures

Discovery services are software products that bring together a library's catalog and citation databases of its choosing. Summon is a discovery service from Serials Solutions. Libraries that subscribe to Summon can choose any number of library resources to be included in their Summon instance, including the library catalog, citation databases. and publisher collections. Serials Solutions builds the Summon index by reindexing scholarly content acquired directly from the publisher, thereby building metadata from the source documents, as well as by ingesting metadata from traditional abstracting and indexing sources. This facilitates the creation of as complete a record as possible for each item and allows Serials Solutions a level of control over the metadata sources used to build their source URLs. The index is continually augmented as matching records are ingested over time: empty metadata fields in the master record are filled in as the information is encountered in other data sources, and conflicting metadata is handled via a formula that generally favors publisher values over third-party data. This continual metadata improvement reduces the "distance" between the original item and the source URL and facilitates continuing improvement of outgoing OpenURL requests from this tool. Because the other discovery tools on the market rely much more heavily on static or externally structured metadata, they lack this advantage.

Unlike the discovery service from EBSCO, Summon contains no native full text, and therefore is entirely dependent on accurate link resolution. As Google's influence continues to reduce users' willingness to search from multiple starting points, the importance of effective discovery tool linking will continue to grow, because of both greater use of these resources and their greater dependence on effective linking. To offer users a competitive alternative to Google Scholar, libraries must implement one-click-to-full-text capability that has a success rate at least as high as Google Scholar's links have. One-click functionality in a results list should work at least as often as links to documents in Google Scholar do. These success rates will vary among libraries because of variation in the effectiveness of their resolver implementations and because of differences in the ratio of publisher-hosted to aggregated content. Google Scholar will have a higher direct link success rate at libraries that license a lot of direct-from-the-publisher full text, whereas Scholar is still dependent on the link resolver to access aggregated full text. Overall, we expect this will result in a renewed investment in link resolver optimization by Serials Solutions, potentially motivating other link resolver vendors that offer discovery products to increase attention to their resolver success rates as well.

Summon
www.serialssolutions.com/summon

Making OpenURL Better: Data, Data, and More Data

OpenURL link resolvers have become a vital part of many libraries' offerings, especially those of academic libraries. As resolvers have become more important, they have undergone the same iterative usability testing and interface improvements that are common for library websites and catalogs. See chapter 2 for suggested improvements in interface design for resolver menus that will ultimately improve the online library research experience.

Only recently has effort been devoted to improving the functionality of resolvers by examining in detail the accuracy of the data that drive them. Also of critical importance is how the standard is implemented within the source databases from which OpenURLs originate. The solutions to OpenURL failures vary widely from library to library and depend on local citation database use and the scope of each library's collection. Improving the resolver at a library that licenses many custom electronic journal packages directly from publishers might require a different approach than would a library that relies more heavily on aggregated databases for full text.

In "The Myths and Realities of SFX in Academic Libraries," published in *The Journal of Academic Librarianship*,[4] the authors summarized user expectations of Ex Libris's SFX resolver, with an eye toward exploring librarians' opinions of the service as well as the impact of this system on the user experience. The authors, librarians at two California State University campuses, analyzed data gathered in an online survey and in-person focus group. They compared these findings with those garnered by analyzing SFX use statistics and test searches. They found the most important issue for users to be the availability of full text articles, while librarians were more concerned with the accuracy of results. The librarians' confidence in SFX was negatively impacted by this concern: they often felt the need to double-check the results by searching a citation database or the library catalog. The article concluded with the statement that user expectations were "slightly higher than" the statistics showed their experiences to be.[5] Causes of linking failures include inaccurate holdings data, absence of selected articles in a target database, or incorrectly generated OpenURLs from a source database. These categories are useful in understanding the inner workings of SFX, but the authors did not analyze their data more deeply to identify the exact causes of errors in each category or where the responsibility for these causes lies.

Industry Initiatives

In 2008, NISO and the United Kingdom Serials Group (UKSG) launched a joint working group charged with creating a set of best practices to address specific problems identified in the UKSG report "Link Resolvers and the Serials Supply Chain."[6] The group, dubbed KBART (Knowledge Bases and Related Tools) published its "Phase I Recommended Practice" document in January 2010, aimed at assisting content providers in improving the serials holdings data that they supply to link resolver vendors.[7] This document contains an excellent summary of the OpenURL process and format specifications that knowledge base supply chain stakeholders can employ for the consistent exchange of metadata. Stakeholders include publishers, aggregators, subscription agents, link resolver vendors, consortia, and libraries. Phase II of KBART's work will expand the data exchange format to encompass e-books and conference proceedings, actively seek publisher endorsement and adoption of the best practices, and create a registry and clearinghouse for KBART formatted data files. See chapter 5 for links to all these resources.

In the final report of a 2009 Mellon planning grant, Adam Chandler of Cornell University investigated the feasibility of a fully automated OpenURL evaluation tool.[8] He recommends that librarians, publishers, NISO and OCLC develop this tool jointly. Such a tool would fill "a critical gap in the OpenURL protocol: objective, empirical and transparent feedback [on OpenURL quality] for supply chain participants."[9] To this end, Chandler proposes that libraries work with vendors to analyze OpenURLs created in source databases, identifying the elements required for successful linking and the frequency with which those elements appear. This analysis of OpenURLs sent from a source database to a link resolver could increase the rate of successful linking. In 2009, a NISO workgroup was created that will build on this work.[10] The Improving OpenURL Through Analytics (IOTA) project is devising and testing a program to analyze libraries' source URLs so that vendors can improve the metadata they are sending to resolvers.

The two initiatives described above primarily address the early steps in the OpenURL process, the building of the knowledge base and source URL processing. A piece not yet addressed is the standardization and quality of how target URLs are parsed by target databases. This is inarguably the least standardized component in the link resolution chain and deserves a similar or greater level of attention than the preceding elements. If more publisher platforms were configured to support incoming links that conform to the OpenURL standard, we could expect to see a significant improvement in target link success rates. Combining an indicator of a publisher's ability to accept standard target URL syntax with the KBART publisher registry would be a significant first step.[11]

Conclusion

The notion of "appropriate copy" is no longer limited to library-licensed content but has expanded to include the

Web. It is impossible for a library to track freely available items on the open Web through its link resolver's knowledge base. OpenURL is still a vital component in the library toolbox, and now that it is a stable and staple technology, industry effort is being devoted to eliminating errors in resolving by examining and setting baselines for the data that drive them. Librarians can play a role in this industrywide effort by looking closely at the efficacy and usability of local resolvers and discovery tools.

Notes

1. Jill E. Grogg, "Linking and the OpenURL," *Library Technology Reports* 42, no. 1 (Jan./Feb. 2006).
2. Beit-Arie 2001.
3. Lorcan Dempsey, "Top Technology Trends" (panel presentation, American Library Association Annual Conference. Washington, DC, June 27, 2010).
4. Jina Choi Wakimoto, David S. Walker, and Katherine S. Dabbour, "The Myths and Realities of SFX in Academic Libraries," *The Journal of Academic Librarianship* 32, no. 2 (March 2006): 127–136, ISSN 0099–1333, DOI: 10.1016/j.acalib.2005.12.008, http://bit.ly/JALwakimoto (accessed Aug. 4, 2010).
5. Ibid., 134.
6. James Culling, *Link Resolvers and the Serials Supply Chain*, Final Project Report for UKSG (Oxford, UK: Scholarly Information Strategies, 2007), www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf (accessed July 31, 2010).
7. NISO/UKSG KBART Working Group, "KBART Phase I Best Practices," Jan. 2010,. NISO website, www.niso.org/publications/rp/RP-2010-09.pdf (accessed July 30, 2010).
8. Adam Chandler, "Results of *L'Année Philologique Online* OpenURL Quality Investigation," Mellon Planning Grant Final Report, Feb. 2009, http://bit.ly/chandler-mellon (accessed Aug. 4, 2010).
9. Ibid., 6.
10. National Information Standards Organization, "IOTA: Improving OpenURLs Through Analytics: Group to Conduct Two-Year Project to Evaluate Metrics," NISO website, www.niso.org/workrooms/openurlquality (accessed July 30, 2010).
11. Adam Chandler, personal communication to the authors, May 13, 2010.