Library Faculty and Staff Papers and Presentations — EKU Libraries

10-1-2010

Rethinking Linking: Breathing New Life into OpenURL

Cindi Trainor
Eastern Kentucky University, cindiann@gmail.com

Jason S. Price is the Collections and Acquisitions Manager at the Claremont Colleges Library <http://libraries.claremont.edu> . He has a PhD in Plant Evolutionary Ecology from Indiana University Bloomington where he cut his teeth as a teacher and researcher before earning an MLS from IU-SLIS <http://www.slis.indiana.edu/> . After spending ten years as a graduate student, he thoroughly enjoys applying his hard won analytical skills to current library challenges. His role as E-resource Package Analyst/Consultant for the Statewide California Electronic Library Consortium provides opportunities to work with publishers, vendors and libraries to improve products and increase pricing equity. He wishes to thank his colleagues on the OpenURL Evaluation team, who collected data for Chapter 3, and especially his family and co-author for their forbearance throughout this ambitious project.

Cindi Trainor is the Coordinator for Library Technology and Data Services at Eastern Kentucky University Libraries, where together with her awesome staff she plans for, implements, maintains and assesses technology in the libraries. She is the former Director of Library/Information Technology for the Libraries of the Claremont Colleges and spent several years at the University of Kentucky Libraries. She is active in LITA and a proud member of the library geek community. She also writes and shoots photos for ALA's TechSource blog, is a co-author of The Darien Statements on the Library and Librarians, and is a photographer whose portraits have appeared in Library Journal, Digitale Bibliotheek and the New York Times. She wishes to thank her colleague, Cristina Tofan, for thoughtful feedback and moral support during the editing process.

Abstract:

Rethinking Linking: Breathing New Life into OpenURL

In this issue of Library Technology Reports, authors Cindi Trainor and Jason Price revisit OpenURL and library linking. The OpenURL framework for context-sensitive linking has been in use for a decade, during which library collections and users' behaviors have undergone radical change. This report examines how libraries can make use of web usability principles and data analysis to improve their local resolver installations and looks to the wider web for what the future of this integral library technology might hold.

Chapter One - Introduction

Scope of this report

The January/February 2006 issue of *Library Technology Reports* 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 (Caplan 2001). An OpenURL link resolver is a software product that takes advantage of this standard to link a citation in one product to that 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 software 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 improving 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:

- the resolver's main purpose to "shorten the path" between citation and item (Dempsey 2010);
- 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 adapted 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 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 SFX customer, an ExLibris 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, freely available online. 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 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. (Fig. 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

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 set up 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 the library's holdings. A knowledge base is generally maintained by 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 back files were purchased. Library staff typically add and maintain holdings data for individual and packages of journals and e-books, 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 the product 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 "back files" 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 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 – the items listed in the resolver menu: native full text from a different source; publisher or electronic journal collection websites; the library catalog; Ask-a-Librarian, Google, etc.

 The OpenURL Process (see Figure 1)

- 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 will present 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 links as a safeguard against malformed or mistranslated article-level links.
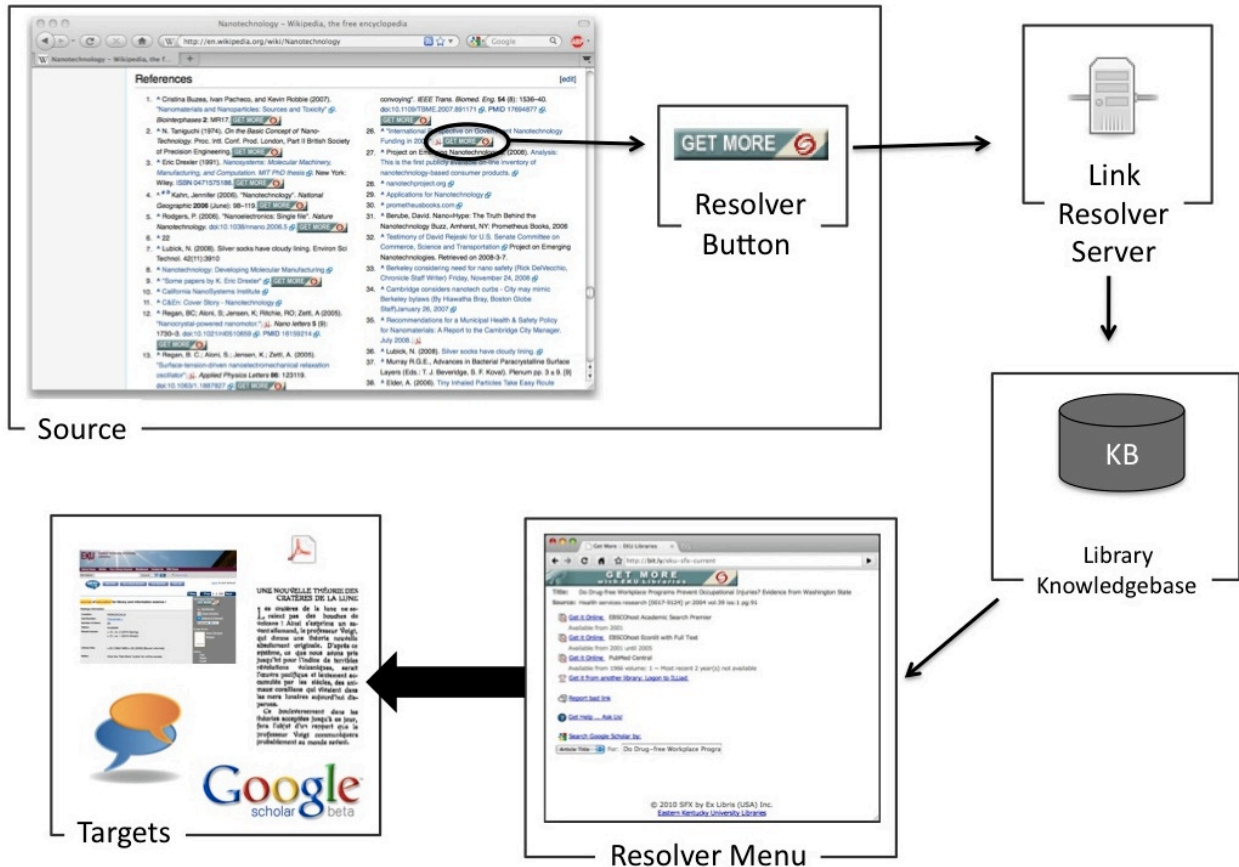
[Figure 1 - How OpenURL works] PDF image courtesy
http://www.famfamfam.com/lab/icons/silk;
Wikipedia contributors, "Nanotechnology," *Wikipedia, The Free Encyclopedia,*
http://en.wikipedia.org/w/index.php?title=Nanotechnology&oldid=374711821
(accessed July 26, 2010).

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 libary collections. However, as pre-prints, 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" today is made more complex 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 (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 like figures, illustrations and tables are 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, 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 fuller explanation of each.

Tapping into the power of Google Scholar

Resolver knowledge bases reflect title level holdings for journals and books but cannot necessarily ascertain whether individual articles are held. As such, 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 post-prints in institutional repositories or random 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 re-indexing 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 source 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 they are 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, both because of greater use of these resources and their greater dependence on effective linking. To offer libraries 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 due to differences in the 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 (including ExLibris) to increase attention to their resolver success rates as well.

Making OpenURL better: data, data, data

OpenURL link resolvers have become a vital part of many libraries' offerings, especially 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 in libraries 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 drives 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 on aggregated databases for full text.

In "The Myths and Realities of SFX in Academic Libraries," published in *The Journal of Academic Librarianship*, (Wakimoto, Walker and Dabbour 2006, 127) the authors summarized user expectations of Ex-Libris' 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" (p 134) the statistics showed their experiences to be. Causes of linking failures included 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 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." 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 companies.   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 (Chandler 2009), Adam Chandler of Cornell University investigated the feasibility of a fully-automated OpenURL evaluation tool. 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" (p6). 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. The Improving OpenURL Through Analytics group (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 in unarguably 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 (A. Chandler, pers. comm.).

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 their link resolver's knowledge base. OpenURL is still of vital importance in the library toolbox, and not 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 drives them. Librarians can play a role in this industry-wide effort by looking closely at the efficacy and usability of local resolvers and discovery tools.

References

Beit-Arie, Oren, Priscilla Caplan, Miriam Blake, Dale Flecker, Tim Ingoldsby, Laurence Lannom, William Mischo, Edward Pentz, Sally Rogers, and Herbert Van de Sompel. "Linking to the Appropriate Copy: Report of a DOI-Based Prototype." D-Lib Magazine, 7:9. 2001. [http://www.dlib.org/dlib/september01/caplan/09caplan.html] Accessed 12/3/2003; URL verified 7/28/10.

Chandler, Adam. "Results of L'Annee Philologique Online OpenURL Quality Investigation."
http://bit.ly/chandler-mellon (accessed 7/29/2010)

Dempsey, Lorcan. 2010. "Top Technology Trends." Panel presentation at American Library Association Annual Conference 2010. June 27, 2010, Washington, DC.

Jina Choi Wakimoto, David S. Walker, Katherine S. Dabbour, The Myths and Realities of SFX in Academic Libraries, *The Journal of Academic Librarianship*, Volume 32, Issue 2, March 2006, Pages 127-136, ISSN 0099-1333, DOI: 10.1016/j.acalib.2005.12.008. (http://bit.ly/JALwakimoto)

CHAPTER 2 - IMPROVING THE RESOLVER MENU: THE MOST BANG FOR YOUR BUCK

Any library can benefit from thinking critically about the use and usability of its link resolver. Many improvements can be made to the resolver interface by applying basic web usability principles; other improvements can be made using tools and reports contained within the resolver itself.

Usability is key to a satisfactory patron experience, whether planning a new OpenURL link resolver installation or seeking to improve a current implementation. It is important to set up a resolver menu so that the sometimes-complex steps to obtaining an item are as simple as possible. Use brief and active language without library jargon, such as "Get it online," rather than "Download full text." If available, take advantage of rules that will minimize dead ends, such as suppressing a link into the library catalog if there are no print holdings, suppressing a document delivery link when online full text is available, and suppressing a link back into the originating database.

Resolver Menu Redesign at EKU

The SFX Working Group was established as a subcommittee of the Online User Experience Team (UX), which was given the responsibility of thinking more holistically about all library web content and systems and improving the usability and functionality of each. Thus, changes to the resolver menu were governed by basic web usability principles. In a redesign process during the summer of 2009, the SFX Working Group at EKU Libraries was charged with analyzing and improving the SFX interface and made the following improvements in our link resolver menu.

At the start of the redesign process, the menu looked like Figure 2:

[Fig 2: former EKU SFX menu; this is not online]

The process of obtaining the text of an item can involve several steps. It can be held by the library online or in print; it can be obtained via Interlibrary Loan; and an increasing number of articles are freely available online, thanks to the Open Access movement. Figure 2 illustrates two of these steps, a link to full text online and a link to the library catalog. A third step, the Interlibrary Loan request, is not present because the article is available online. The thinking at the time--whether deliberate or not--mirrored the library instruction process by which undergraduates were introduced to online searching. This numbering system and the extra text it represented were removed in our redesign. The Interlibrary Loan request was also added to every menu, facilitating user request of articles from the menu that cannot be found or that are falsely represented in the knowledgebase. We significantly reduced the amount of text used to describe each service and collapsed the listing for each service into a single line, with a second line for any available holdings. Outdated "Go" buttons were

removed in favor of linking the action: "Get it online," "Get it in print," "Get it from another library." Small icons were added as visual cues relative to each service.

We also created additional services, integrating more options into the menu. Figure 3 illustrates the redesigned menu with online full text targets, and Figure 4 illustrates the redesigned menu with print holdings and the Distance Education Request, one of these additional services.

[Fig 3: current EKU SFX menu; link: http://bit.ly/eku-sfx-current]

[Fig 4. EKU SFX menu with "Get it in Print" and "Distance Education Request" links; http://bit.ly/EKU-print]

"Report Bad Link" - The link to any menu can be sent to the Electronic Resources librarians with a single click. Users are required to leave their name and email address and are given the option of leaving comments. A link to the menu is sent via email to EKU's Electronic Resources Librarians, who frequently respond to link reports within one to two business days. This target was used 225 times in its first year, or only one-tenth of one percent of the time that it appeared on menus, an interesting statistic in itself.

"Search Google Scholar" - This target was used nearly 500 times the first two months it was available, or 6% of the time that it appeared on SFX menus. As our testing revealed, this target is particularly useful for items available via Google Books and for articles and conference proceedings available via open access from publishers or in institutional repositories. Previews, tables of contents, reviews, tags, maps, and other information are often available for books. A Google Book Preview often provides enough text to indicate to the user whether obtaining the book via Interlibrary Loan would meet his or her needs.

"Distance Education Request" (see Fig 4) - EKU has four regional campuses and other centers located in the University's 22-county service region and an increasing number of online-only students located nationwide. EKU provides equivalent library services to these students as defined by the ACRL Guidelines for Distance Learning Library Services, including mediated access to print materials located on the main Richmond Campus. The "Distance Education Request" button in SFX offers a quick and accurate way to submit these requests. It is much preferable to the former method, which required users to copy and paste each field of the citation, as well as their personal information, for each request. The new request button was used more than 80 times in its first month, during a summer semester, and increase of more than 10 fold above the same time the previous year. Discussions are underway at EKU for combining and streamlining our current three delivery services to eliminate needless referrals and to deliver as many requests online as possible, rather than via ground courier. Those services are traditional Interlibrary Loan, Document Delivery, and Distance Education requests. It is expected that SFX and Illiad will play significant roles in this transition.

"Get it in Print" (see Fig 4). Print holdings were activated in our SFX knowledgebase, but we did not add holdings. When SFX finds an ISSN match for a journal, a target link to the catalog is presented. It is important to note that because this is a match at the title level rather than at the issue level, this method occasionally results in false positives, such as when the library is missing issues or does not own a complete print run. There are several options for making print holdings information available in a resolver menu. In the future, EKU may revisit David Walker's Chameleon SFX plug-in to enable real-time look-up of book and journal holdings, but the journal holdings in the catalog must be cleaned up and made consistent before this can happen.

Resolver rules and direct-to-full-text

Resolver products typically have rules built in for conditionally displaying a target, taking advantage of CrossRef DOI linking, and pushing the user directly to the full text of an article instead of displaying the resolver menu.   Examples of conditional target display include suppressing a link to the Interlibrary Loan form if full text or print holdings are found and suppressing a link into the database from which the openURL originated. Before enabling direct links to full text in lieu of displaying the resolver menu, a library must test this capability and exclude any targets that don't work reliably. A library could conversely enable only those targets that are most highly-used or that resolve most reliably. It is important that these full text windows display a button or banner that can return the user to the full resolver menu. If for some reason the target item is not available, the menu will facilitate the use of other services, rather than being a dead end.

Use Statistics: SFX

Examination of reports built into the link resolver can assist libraries in deciding where to focus energies when seeking to make improvements. SFX makes several statistical reports accessible that reveal how the system is (and isn't) used. While many of the reports, such as "Journals Requested Without Full Text" and "Unused Full Text Journals" are geared toward collection management, some can be used to examine resolver functionality. For an excellent introduction to the queries available in a standard SFX installation, see the Chrzastowski, Norman and Miller article, the full citation for which is available in Chapter 5.

Using these standard SFX queries, it is possible to gauge what sources and targets are used the most, which could assist in identifying priorities for testing. The "Top Target Services Shown in the SFX Menu" (Query 6) reveals the number of times a particular target and its concomitant service have been displayed in SFX menus over a period of time. To get an idea of how these requested menus were used, the report from Query 6 must be combined with Query 7 or Query 8, which detail "click-throughs," or how many times each target was clicked by users.

EKU generally has only one service per target. Libraries that use more than one service per target (for example, those which enable GetFullText *and* GetAbstract or GetAuthor) could use Query 8, "Number of Click-Throughs per Target Service," to get an idea of the demand for these different services. In Table 1, results from Query 6, "Top Target Services Shown in the SFX Menu," were combined with Query 7, "Number of Click-Throughs per Target," to begin to paint a picture of high-demand targets. The table is sorted by number of click-throughs.

Table 1: Top Ten Click-Through Targets at EKU, April 2010

Target	Requests	Click-throughs	Click-through Rate
Interlibrary Loan Request	40039	1282	3.20%
EBSCOHost Academic Search Premier	1663	889	53.46%
Library Catalog	6545	763	11.66%
Miscellaneous Free Ejournals*	1351	557	41.23%
Elsevier Science Direct	513	385	75.05%
EBSCOHost CINAHL with Full Text	517	324	62.67%
EBSCOHost Business Source Premier	559	275	49.19%
Miscellaneous EJournals**	448	269	60.04%
Sage Criminology Full Text Collection	305	245	80.33%
Gale Opposing Viewpoints	474	184	38.82%

*Miscellaneous Free Ejournals is a target comprising 18,302 individual titles at EKU that are freely available via the web. These websites are not available on other platforms. These titles are not sent through the library proxy server.

**Miscellaneous EJournals is a target comprising 95 individual journal titles at EKU that are not available on another target platform. These are proxied titles.

Table 1 provides a good indication of where EKU could focus our energies in testing target URLs. Errors in highly-used targets affect more people, and therefore fixing these errors would benefit the highest number of users.

Query 19, "Most Popular Journals," displays a title and ISSN list of the most frequently-used journals linked from the resolver, as well as the number of times the title was presented in a menu and subsequently clicked. This query's results could also be used to prioritize target links for testing. Table 2 shows the top five most popular journals at EKU in April 2010.

Table 2: Top Five Journals at EKU, April 2010

Journal	Requests	Clickthroughs	Clickthrough Rate
Dissertation abstracts international	97	21	21.65%
Dissertation abstracts international. B, The sciences and engineering	96	13	13.54%
Science	77	73	94.81%
Criminology	68	37	54.41%
Journal of Criminal Justice	67	52	77.61%

These results are particularly disturbing, as the top two "journals" are in fact requests for individual dissertation titles that we know without exception failed. Upon further examination, we find the source URLs to be technically correct. These titles may very well be found in ProQuest's Dissertations Full Text, but the resolver is not able to translate citations from Dissertation Abstracts International into links for items within the Dissertations Full Text database. See Chapter 3 for an explanation of the workaround needed to fix this problem.

The query that is perhaps the most useful for troubleshooting individual journal titles is Query 20, "OpenURLs that resulted in no full text services, selected by source." This query displays individual journal titles used by patrons but for which no full text targets are presented. Full text for articles can be unavailable for varying reasons: the article in question lies within an

embargo period; the library's online subscription does not start early enough; the online version contains only selected full text. The full text of books, chapters, dissertations and other formats is sometimes not found due to errors in the OpenURL syntax; see Chapter 3 for examples of how to examine and code source OpenURLs like those identified in this query. Query 20 results are listed by source; a source that lists many OpenURLs in this report might be a good place to begin troubleshooting. Source URL troubleshooting requires communication with database vendors and publishers. It's worth noting that administrators of locally-hosted SFX installations have ability to edit source parsers, the source-to-resolver translators; this facilitates addressing persistent source problems locally.

SFX Query 11, "Most Popular Journals Selected by Source" can be used to list the journals used most frequently for any given source database. These reports might be used to identify how a new database is performing, or to estimate how widely word of a trial database spread across the community.

Caveats

The SFX Queries module is not intuitive to use. Even with the excellent primer found in Chrzastowski, Norman and Miller, navigating the interface can be difficult. We suggest the above queries as a starting point for a discussion among library staff about which sources seem most difficult to use and which targets seem to fail most frequently.

Codes for source databases are nearly unfathomable, making the reports that are generated by source or which present results sorted by source are particularly difficult to interpret. A key to interpreting the Source ID, or the "sid," would be helpful to the library community but does not yet exist.

Serials Solutions does not provide standard reports that address resolver usage except at the level of the total number of click-throughs for a given time period. This number of click-throughs is compared with the A-Z title list and with 856 links clicked in the library catalog, for customers that use MARC records generated by Serials Solutions. Target click-to statistics are available, but they do not separate A-Z list or MARC record use. OpenURL server logs that could be parsed by customers are not readily available, though they can be requested. We hope that Serials Solutions will invest development resources in making 360link evaluation possible as a part of their core assessment utilities.

Chapter 3

Abstract

OpenUrl link resolvers have become a core component of a library user's toolkit, yet a historical comparison suggests that they fail nearly a third of the time, and have not improved over the past six years (see Table 3). This study dissects the evidence of failure types and causes for two resolver installations in order to identify and prioritize specific tasks that libraries can undertake to accomplish incremental improvements in their resolver's performance. In doing so, we hope to stimulate understanding, thinking, and action that will greatly improve the user experience for this vital tool.

======

The preceding chapters of this report address the state of the art of OpenURL (Chapter 1), and general improvements that libraries can make to their local link resolver implementations (Chapter 2). This chapter reports the results of a detailed study carried out to determine link resolver accuracy rates and to tease out the causes of link resolver failure at the authors' institutions[i]. In addition to quantitative assessment of local resolver functionality, we gained valuable qualitative experience as extensive users of our own systems. The results of these two types of observation are then combined into a top ten list of tasks that should accomplish significant improvements in link resolver effectiveness. The majority of these tasks are broadly applicable, and many can be applied individually to improve resolver effectiveness at any library.

[[Table 1 Number of citations (= source OpenURLs) tested by database and document type]]

Testing OpenURL full text link resolution accuracy at our institutions

This study is based on the "real-life" approach of Wakimoto et al. (2006) to allow a historical comparison with their 2004 SFX testing results. Resolver results from likely keyword searches

for a number of popular databases were tested from September 2009 through June 2010. Stratification by document type was added to increase exposure of non-journal resources. Each author tested seven databases, collecting results for journal articles (10), book chapters (5), books (5), dissertations (5) and newspaper articles (5) whenever citations to those document types were available in the source database (Table 1). Citations that include native full text were avoided, as well as those from journals or books that had been tested previously.

[[Table 2. Number and proportion of menus with full text links offered by each institution]]

Overall, 351 source URLs were tested in this study. About half of the resulting resolver menus offered one or more online full text links (n=169 [48%]; average full text link number = 2.01). The other half of the menus indicated that no full text was available, offering links to search the catalog, populate an ILL request and search Google Scholar instead (Table 2). Every full text link was checked for access (n=343), and Google Scholar and Google were searched for each result with no full text available (n=182). The results were then coded into six categories, mirroring Wakimoto et al.'s (2006) designations. Their results are included for comparison (Table 3).

[[Table 3. Resolver results for full text requests in each dataset (after Wakimoto et al. 2006)]]

Wakimoto et al. reported that about 20% of their resolver results were erroneous. Roughly half of the errors incorrectly indicated availability (false positives), while the other half incorrectly failed to indicate availability (false negatives). Our result rates for these errors were similar. However, the 'Required search or browse for full text' category was re-assigned from the 'Correct' group to the 'Error' group to reflect reduced user willingness to search or browse for the full text. When the target full text item or abstract with full text links is not presented on the target page, most users and even many librarians perceive the resolver as having failed. This category increases the total error rate by nearly 70%, averaged across both datasets. This results in total error rates of 35% for the Wakimoto, et al. dataset and 29% for our dataset (Table 3).

The error rates increase further when freely available content is taken into account. All "no FT available" items were searched in Google Scholar and Google, using links provided from the resolver window or with the LibX browser add-on. Twenty-one of 138 (15%) were available via the web. Tapping into this content is equivalent to increasing our budgets by 15%. Furthermore, the percentage of 'externally available' items is likely to be higher in an article-heavy dataset, and will increase over time as authors continue to post their own content on personal web pages and in institutional repositories. This additional category of false negatives increases the overall error rate to (33%). While expanding resolver knowledge bases to enable direct retrieval of 'external' items may not currently be possible, we can accomplish improved access to them from our resolver windows. As a first step, links to extend full text retrieval to Google should be made more prominent in resolver menus. It should be our eventual goal to fetch the full text link (or even the document[ii]) from the web and present it in the resolver window.

To be fair, there is a less critical way to measure resolver success: how many resolver menus that offer full text contain at least one link that leads directly to accessible full text? By this definition, the CUC resolver was successful 93% of the time (in 86 of 93 menus), and the EKU resolver was successful 70% of the time (in 54 of 77 menus). Thus, by this measure, the resolver was successful approximately eight out of ten times for the combined dataset.

Resolver result accuracy by document type

The opposite of the resolver error rate is the accuracy rate: 71% overall for the citations tested. Book chapters and book menus were far more accurate than those for other document types (0.98 and 0.95, respectively, Table 4). Unfortunately, the vast majority of these successes (101 of 105) correctly reported no online access, reflecting small e-book collections and/or their absence from the knowledge base. In addition, because the study was designed to emphasize book content (40% of the source URLs tested), the overall accuracy rate is probably an overestimate of what most users experience. Indeed, when book results are excluded, the overall accuracy rate is reduced to 64% (270 of 420 results). With this in mind, our results suggest that only about two out of three resolver results are accurate.

[[Table 4. Resolver full text link accuracy rate by document type and source database]]

In contrast to book content, newspaper and dissertation results had much lower accuracy rates than average (0.38 and 0.30, respectively, Table 4). Newspaper article citations occurred in only two of the databases, and yielded contrasting accuracy rates. EBSCO's Academic Search Premier citations had many more bad links than Serials Solutions' Summon. This is probably at least partly due to the restricted newspaper content in ASP: the Wall Street Journal and New York Times are notoriously hard to link to. It is also possible that Summon's unified index has improved the success rate for this document type. More data is necessary to distinguish among these alternatives. In contrast, the data for dissertations is quite consistent. Accuracy rates were very low across the board, with most of the successes attributable to specialized indexing (as in Summon and ERIC), or to older results that were correct by default because full text is not available online. We further address the poor accuracy rates for newspaper and dissertation content in the section on error causes, below.

Nearly two-thirds of the results were for journal articles, so perhaps not surprisingly, their accuracy rate most closely mirrored the overall results (75%, Table 4). America: History and life (AH&L) had an unusually high success rate (0.97), while the National Criminal Justice Reference Service (NCJRS) database was on the very low end (0.18). The high error rate for NCJRS is attributable mostly to the limited metadata sent in its source URLs. They include only journal title, date and article title. The reason for AH&L's high success rate is less clear.

Although it is tempting to further analyze our accuracy results by source database, we deliberately chose not to do so, for three reasons. First and foremost, although source URL quality can influence linking accuracy, they are the furthest from the final result, being dependent on the 'downstream' resolver and target database. Secondly, only one document type could be tested across all citation databases, and half of the database/vendor combinations were only tested at one institution. Finally, the IOTA project (Improving OpenURLs through Analytics, http://www.openurlquality.org/) is focused on assessing source URL quality for large OpenURL datasets, and is better positioned to do so. Instead, we present an analysis of the *causes* of failure recorded in our study. To our knowledge, this is the first systematic attempt to categorize the causes of a set of OpenURL failures and determine their relative frequencies. It is our hope that these results will help determine which aspects of the

resolution chain need the most attention and identify solutions that will address the most common failures.

Failure cause analysis procedure

Librarians and OpenURL aficionados alike often disagree as to who or what is at fault for link resolution failure. Some say it is poor standards implementation or metadata quality in source databases. Others blame their link resolver vendor and advocate for switching to a different supplier. Still others claim that it is poor holdings data in the library's knowledge base. The final scapegoat is the full text provider, which may fail to resolve perfectly-formed (and standardized) target URLs. In one sense, the answer is simple: each component contributes to the problem at least some of the time. But this simple answer reveals the key question: which component or components are most commonly at fault in any given library? It remains to be seen whether generalizations can be made. It is certainly true, however, that for particular combinations of source, resolver, knowledge base and target, some components are more at fault than others. Libraries should evaluate and improve these components for their most important sources and targets. This section presents the framework of a rubric which can be used to do so.

Analysis of the causes of OpenURL link resolution failure is inherently a step-by-step process, although upstream errors can often be corrected by downstream components. For example, missing or inaccurate journal title data in a source URL can be added or replaced by a resolver that maps ISSNs to journal titles. Similarly, conflicting data in a target URL can be surmounted by a full text provider algorithm that accomplishes linking from a subset of the metadata elements that *do* match an item available from the provider.

In order to identify the cause of each resolver failure, a wide range of data for was collected for each full text resolver result [see Exhibit A: refer to copy of the MS Excel workbook archived by ALA and/or the CCDL]. These included the source URL link to the resolver menu, the resolver results details (including the outgoing full text link and resulting provider target URL, where applicable), the nature of the result set a the target, and notes to explain the result, as necessary. Finally, in each case where full text could not be accessed through links

in the resolver menu, we checked for full text availability at the provider site and elsewhere on the Web.

Failure cause by error type

In general, the causes of resolver failure were evenly distributed across the OpenURL resolver chain. No more than 20% fell into any of the eight categories (Table 5), and no more than 33% were due to any of the five components (delineated by bold lines in Table 5). In fact, when the 28 resolver translation errors that were due to dissertation citations (see Table 7) were dropped from the analysis, no single component was responsible for more than 26% of the errors. Despite this even distribution of causes, some interesting general patterns emerge, particularly when they are viewed by vendor/database and document type.

[[Table 5. Frequency of failure causes by error type]]

It is important to note here, however, that there are two cause categories that could not be assigned to one of the three true resolution components (i.e. data source, resolver or provider). This is obviously the case for the miscellaneous category, by its very nature. However, 9 of the 15 'miscellaneous' failures were due to CrossRef errors in CUC's resolver, which weren't analyzed further because they are external to the normal OpenURL resolution chain and beyond the control of 360 Link customers. The second category is more troublesome. Twenty-one of the errors which required search or browse could not be distinguished as the responsibility of the resolver vs. the provider. This limitation is inherent in the translation specificity of the target URL for a number of providers: was the search/browse required because (1) the target URL didn't contain the data necessary for item-level resolution or (2) item-level resolution is not supported by that particular provider? Item-level resolution in NewsBank is a likely example of the first case, since making changes to the target URL can send the article title to its native search. The Directory of Open Access Journals is an example of the second case, since it represents an 'aggregated provider' where different journal websites vary in their ability or syntax to support deep linking. Thus this category is a particular challenge for the resolution chain, but should also represent fertile ground for improvement of linking to particular high priority providers. These improvements

can be accomplished by fixing the translator where this is possible (case 1) or by replacing the journal-level link with an item-level link to search Google Scholar (Case 2).

[[Table 6. Frequency of failure causes by source Vendor/Database]]

Failure cause by vendor and database

Interesting patterns are revealed when the failure causes are analyzed by vendor and database. For source data quality at the vendor level, EBSCO and Serials Solutions outperformed, while CSA, Google and OCLC contained all the errors (Table 6). Despite its wide universe of source data, the Serials Solutions Summon source data tested was error-free, perhaps a testament to the success of their 'unified index' techniques. EBSCO's tested source data was also free of errors, despite the dual institution sample for 3 of the 4 EBSCOHost databases tested. This is likely due to a combination of high-quality indexing in Academic Search Premier (ASP) and the particular databases tested on this platform. CSA's failures were restricted to two of the four Illumina-hosted databases. Most of the errors derived from an externally-produced index (National Criminal Justice Reference Service-NCJRS), although some came from a database for which CSA took over indexing in 1999 (Sociological Abstracts-SocAbs). The CSA results lend credence to the perception that source databases vary widely in their source URL quality[iii]. It is not surprising that Google Scholar had a number of source URL errors, given its crawler-based indexing approach[iv]. The high ratio of source errors from the results tested from OCLC Worldcat.org (from a single institution) may reflect lower quality indexing in by ArticleFirst (produced by OCLC since 1990), Worldcat.org's disparate sources of index metadata, or the nature of the journals in the discipline chosen for the search. On that note, it is important to add a caveat to the preceding discussion.

Because we did not control for variation in search topic, publication date, or total number of citations tested from the various vendors and databases (and these are just a few of the potentially confounding factors), the speculation in the preceding paragraph should be viewed with an especially skeptical lens. That said, there are few, if any, other patterns that emerge from this level of analysis. Twenty-three of the 33 errors (70%) that were attributed to the provider component occurred for citations from Academic Search Premier or Summon, but these can hardly be blamed on the source, particularly with their spotless source URL record.

Furthermore, nearly two-thirds of these errors were for newspaper articles, and are probably largely attributable to the vagaries of this document type.

Failure cause by document type

The last level of failure cause analysis examines their relationship to document type. Particular categories of failure were much more common in citations of one document type than in others. Recognizing these differences can help to identify which aspects of the OpenURL resolver chain need the most attention for dissertations, newspaper articles and journal articles.

[[Table 7. Frequency of failure cause by document type]]

Dissertations provide the best example because two error categories were clearly over-represented for this document type: resolver translation errors and source URL inaccuracies (Table 7). Of the 60 dissertations tested (42 of which failed), nearly half of them failed to link to full text that is available from Proquest's Digital Dissertations due to a resolver translation error. To rectify this situation, both ExLibris' SFX and Serials Solutions 360 Link need to translate post-1996 citations for Dissertation Abstracts International (DAI) into a search for the full text by the dissertation title (atitle) in Digital Dissertations. This should be applied to all genres, but particularly to "genre=article," as most indexes still treat DAI as a journal that a user would want to retrieve articles from, even though it is available only in print and contains only abstracts. It is also common for the genre of a dissertation to be erroneously indicated as 'book' in source URLs. About a quarter of the dissertation failures were caused by this error. In Sociological Abstracts (5 of the 10), these can be resolved by matching the publisher field in the source URL (with Proquest, Ann Arbor MI). Unfortunately, each database provides different clues that these 'books' are dissertations, so distinct solutions are required for citations for each source database. When these errors are universal and consistent within a highly used database, however, it is worthwhile to implement custom fixes. Such efforts bring up a key distinction between the two most popular link resolver vendors. With locally-hosted SFX implementations, it is possible for the library to customize source URL resolution by editing the source parser[v]. For 360 Link, customers need to advocate for a global fix in each specific database. Obviously, each situation has its drawbacks.

Nearly half of the newspaper article resolution errors were due to target URL translation errors (Table 7). This suggests that improved outgoing target URL translators are the most appropriate fix for libraries or link resolver vendors that choose to prioritize increased accuracy for newspaper articles. Although there are many fewer providers of newspaper article full text than of journal full text, accuracy rates for correct resolution of newspapers are apparently still quite a bit lower than for journal articles. Although these errors made up only ~20 % of the errors encountered (28 of 153), they appear to be quite common, since they resulted from only 4% of the citations tested (i.e. 15 newspaper of 350 total source URLs). These figures suggest that the payoff per provider target fix will be greatest for newspaper article providers.

Journal article errors were caused by failures all across the possible spectrum (Table 7). Furthermore, they were quire evenly distributed: at least 16% were attributed to each of the 5 resolver components. These errors were most commonly caused by source URL data problems (23 of 79), with two-thirds of these due to erroneous data and one-third due to missing data. The wide spectrum of causes for journal article full text resolution failures suggests that the best approach for this document type might be a journal-level approach. We recommend that libraries work from a prioritized list of their most-used journal titles.

Qualitative observations on resolver effectiveness

Our study also provided a great deal of insight into the effectiveness of our resolver menus that is not reflected in the data presented above. As active users of the product, we noticed a number of aspects of the front-end functionality that need improvement. These observations pertain to the specifics of OpenURL functionality, providing a complement to the application of general web usability principles to resolver menus in chapter two. We present them below as specific constructive criticism of our own systems, but most will apply to resolver implementations at other libraries.

The primary user expectation when clicking the resolver button is that it will lead them to full text. Given that about that about half of the requests sent to our resolvers do not match full text covered in our knowledge bases, it is important to make these results as clear and

effective as possible. At EKU, the notification states that "this item is not available online" (e.g. http://bit.ly/ekuNotOnline). Although the statement is clear and simple, it is false for items that are accessible on the Web but not represented in the knowledge base (as in this example). At Claremont (CUC), the phrase is "No full text for this citation was found in the online collections of the library." Although technically correct in all cases except for knowledge base errors, this text is wordy and is not the most important information for the user at that point of need. Put another way, the user generally does not *care* whether the item is in the library's collection: they clicked the resolver button because they want to know whether the item is immediately accessible to them. This principle calls for an interface improvement that is far more important than the terminology. We need to restructure our resolver menus so that additional instantaneous paths to the full text are co-located with the results from the knowledge base. Thus we recommend that the links to extend the full text search to Google Scholar be moved up to the second position in the resolver result menu rather than being placed near the bottom as a solution of last resort. This is a particularly important improvement for CUC, whose resolver menu is very long and interjects links to search for related articles above its additional options (e.g. http://bit.ly/cucWbx).

There are also a number of cases where identical target links are presented in the same menu. For example, a "Get It Online" link is presented for a single version of an article that is listed both in EBSCOHost Academic Search Premier and EBSCOHost EconLit with Full Text (see http://bit.ly/eku-sfx-current) or in a publisher site as well as from CrossRef. At best, this adds text to the menu that is not needed when the first link works. At worst, when the first link doesn't work, the user will try the second link, thinking it is different, and that link will fail as well. This usability issue can largely be solved by adjusting the resolvers' administrative settings, although these settings do not affect CrossRef links.

Order of link presentation is a thornier issue. It would improve the user experience to be able to order them by some combination of **link depth,** e.g., article-level vs. journal-level; **format(s) available**, listed in order of preference--HTML+PDF, PDF only, HTML only, HTML lacking figures or tables, and selected full text [i.e. some items missing]); and **reliability**.

- **Link depth** should be consistent within a particular provider, so it is particularly useful to have an administrative choice that would allow demotion of hosts based on this property. This seems particularly important for optimizing 'one-click' or 'direct link'

functionality. When title-level links must be used, it would be extremely valuable to include a banner at the top of the journal homepage with the citation specifics (like WorldCat does, http://bit.ly/dns81H)

- The item **format(s) available** differs between providers, and within providers among titles, and even within single titles. Although this information is certainly known by the provider, it is not commonly shared, and was excluded from a draft list of data elements that KBART considered requiring (see Chapter 1, Industry Initiatives). It seems reasonable to require providers to indicate whether portions of articles and even whole articles are missing for each title, but this too has not been forthcoming, except in extreme circumstances.
- **Reliability** is certainly the most important of these three criteria, but it is also the hardest to measure, presumably because the extent to which target links actually result in full text access is not captured by OpenURL server logs. The Pubget PDF delivery service (Chapter 4) may have unique insight into these numbers.

The resolver menus for book chapters and books at CUC need attention. They are specific to the resource type (genre) for 360 Link customers. Both menu types require a catalog search to determine whether the book is available online; it is far preferable to indicate online availability in the resolver menu. Furthermore, both menus are set up to search the local and union (Inn-Reach) catalog in separate steps (http://bit.ly/CUCbc), even though the local catalog will send the search through to the union catalog when requested. They are also set up with separate target links by ISBN and Book Title, and the ISBN search regularly fails because the resolver adds and then searches by 13-digit ISBN, while the local catalog predominantly contains the 10-digit version. When sending book chapter searches from the CUC resolver menu to Google Scholar, a chapter title is sent, but this does not directly facilitate searching for the book title in Google Books. EKU's Google Scholar search for book content (http://bit.ly/EKUbk) is preferable, although sending searches as phrases, (i.e., in quotation marks) would improve results. Google offers results for keywords when the phrase search produces no results, so nothing is lost by sending the search in this manner[vi].

Top ten list of tasks to improve resolver effectiveness
These tasks are presented roughly in order of increasing complexity. That said, they involve a wide variety of skills, so the degree of challenge of each will depend on the expertise available at each library.

Examine the "no full text link provided" report **(SFX only)** In addition to being a valuable collection development tool, SFX usage report Query 20, "OpenURLs that resulted in no full text services, selected by source," provides an excellent opportunity to test for false negatives (see also Chapter 2). It combines source URLs that fall into the first and last result categories (Table 3), supplying a list of URLs that can be tested for access using Google Scholar links from the corresponding resolver windows. Patterns in this data may reveal whole collections that are not listed in the library knowledge base, a problem that is easily rectified. It is also easy to assess the extent of the requested content that is available on the open web as a part of this process.

Fix Dissertation target linking EKU's usage and OpenURL failure data provide powerful justification to fix linking to this class of resource (See Chapter 2, Table 2, and Table 4, above). Because an improved source parser provided by the link resolver vendor seems to be the ideal solution, we are requesting a global fix of this issue by Serials Solutions and ExLibris. In the meantime, locally-hosted SFX implementations can edit their own source parsers to fix this problem (cite Resources section). Our results showed that newspaper article linking failed almost as often as dissertation linking. Although newspapers are least as significant a concern, their pagination and date variation, short non-distinct article titles, and frequent supplementary sections make them much more of a challenge.

Review every full text provider for item- vs. title-level linking Given the overarching goal of reducing the number of clicks from the resolver button to the full text, item-level "deep linking" is always preferable. In most cases, link level is determined by the target parser, which translates the OpenURL into a request that the full text target platform can process. Obviously, it makes sense to start with the most frequently requested providers, examining them for item- vs. title-level linking and ensuring that successful item-level linking is established whenever possible. Furthermore, knowledge of this attribute is essential for establishing the order in which full text links are presented.

Reorder the full text provider links This is an art rather than a science. It is, nonetheless, very important, because of the tendency of users to click on the first link, and because one-click access is heavily dependent upon it. Key provider factors include

linking level; full text format and completeness; and reliability (discussed above). Once the values for each of these factors are known for each full text provider, the library can decide how to weight each factor. After the most desirable order is determined, it can be integrated into the administrative settings. By default, both systems list targets alphabetically. For 360 Link, setting the order requires entering in a rank order number for each database, not each provider. This leaves a lot to be desired because many providers have multiple databases which should receive the same rank and minor adjustments require extensive re-ranking. Perhaps a simple solution would be for Serials Solutions to change their system to allow priorities (i.e., 1, 2 or 3) rather than a ranking (1 to 314 for CUC), or even to offer their own order, based on the factors above. SFX is significantly simpler to configure: it only requires insertion of the list of targets in the desired order in a configuration file. SFX also provides the ability to force specific targets to appear at the bottom of the list, allowing implementation of a simpler ranking (i.e. 'good' and 'bad').

Expand knowledge base coverage and rework resolver menus to maximize full text access There is a delicate balance between expanding knowledge bases to cover more free and open access full text content and reducing resolver effectiveness, because these resources tend to be less well managed (Hutchens, 2009). A first step here is to maximize use of freely-available collections that are covered by commercial knowledge bases (see data on error rates from Hutchens reported by Brooks-Kieffer, 2009). Libraries can balance more extensive knowledge base coverage with more prominent and effective links to use Google Scholar and Google to access these resources (see section in 'Qualitative results' above).

Another key area of knowledge base expansion is the inclusion of e-books. Although there are rudimentary implementations of these in both vendors' products, there is still a great deal of room for improvement. Since libraries are investing considerable effort representing e-books in their catalogs, the best near-term solution is probably an adaptation of David Walker' Chameleon SFX plugin (http://www.exlibrisgroup.org/display/SFXCC/Chameleon+SFX+Catalog+Integration+Plugin) to integrate e-book lookup into the full text services section. A similar JavaScript-based tool could potentially be built for 360 Link.

Optimize top 100 most requested journals According to the 80/20 rule, 80% of use occurs in 20% of the titles, so focusing on heavily-used journals will address a great deal of the overall usage. Although only SFX provides a report is specific to resolver requests, 360 Link customers can use the core Usage Statistics report "Click-through statistics by Title and ISSN" to list their 100 most popular titles. A general citation database can then be used to test resolution to articles in these journals, allowing libraries to assess the associated success rates and failure causes, as demonstrated in this chapter. When the underlying data are collected in a systematic way, spreadsheet pivot tables can be used both to examine frequencies and to show details from individual categories [Exhibit A]. This transforms the spreadsheet into a rich, easily-accessible archive of examples that can be used for troubleshooting and sharing with others. Although some issues may be beyond reach, many can be addressed successfully, once they are recognized. Priorities can be established based on the frequency of the problems and the relative ease of fixing them.

Optimize top 10 full text target providers The number of click-throughs per target host (Ch. 2, Table 1, SFX Query 7) can be approximated with Serials Solutions "Click-Through Statistics by Title and Database (Holdings)" report.

Extract and harness the resolver use data to better inform a top down approach The most efficient approach to improving the user experience with OpenURL linking requires identification of the fixes that will be of greatest benefit. SFX libraries can gain significant insight into usage patterns via its standard usage reports (see chapter 2 and Chrzastowski 2009). However, the most powerful source of this information is the resolver server log. The structure of the OpenUrl standard makes analytics on these files particularly fruitful. For example, extraction of data for 'sid=' and 'genre=' provide valuable information on the most-used citation databases and content types. Sorting these files by web domain separates source URLs from target URLs and free web analytics software (e.g. Funnelweb http://www.quest.com/funnel-web-analyzer/) can identify their frequencies. Resolver log files will be a crucial source of information for 360 Link customers, who do not have access to resolver reports like those contained in SFX. Regular collection of these files can also support database evaluation and other collection development needs (e.g. Orcutt, 2009).

Optimize top 10 source databases by content type Once staff at a library extracts a list of the frequency of requests by content type for its most-used citation databases from a log file, they can optimize resolution from these key combinations in the manner described above. For example, there may be a high volume of requests for book chapters in PsycInfo or books from MLA. Optimization of alternative content types is likely to include menu reformatting, in addition to the data- and translation-related issues common to journal article resolution. This level of analysis may also reveal peculiarities that are unique to the specific key combinations, thus revealing important issues that wouldn't show up in standard usage reports.

Implement, test and optimize one-click/direct link to full text As noted in Chapter 1, discovery tools will be dependent on one-click if they are to be a viable alternative to Google Scholar. Also, It seems likely to us that in the future, link resolution will be passive and menu-free, rather than active and menu-based (see e.g. Pubget Ch. 4). The first step toward this eventuality is implementation of the one-click to full text service. We chose this as the final recommended step, not because it is the most complex, but because all of the previous improvements will make it more effective. In particular, reordering the full text provider links should be a prerequisite to this. One link resolver feature that is needed here (not yet offered by 360 Link) is the ability to 'opt out' of one-click for source databases and full text providers that are problematic. This function is available in SFX, at least for full text providers.

Notes (see also 6 footnotes, inserted in text)

Jina Choi Wakimoto, David S. Walker, Katherine S. Dabbour, The Myths and Realities of SFX in Academic Libraries, *The Journal of Academic Librarianship*, Volume 32, Issue 2, March 2006, Pages 127-136, ISSN 0099-1333, DOI: 10.1016/j.acalib.2005.12.008. (http://bit.ly/JALwakimoto)

Hutchens, C. Managing free and open access electronic resources. UKSG Serials-eNews. no. 210. 11 Dec 2009.

http://www.ringgold.com/UKSG/si_pd.cfm?AC=0350&Pid=10&Zid=5067&issueno=210

Brooks-Kieffer, 12 Feb 2009. ER&L 2009: Managing free e-resource collections. http://ksulib.typepad.com/conferences/2009/02/erl-2009-managing-free-eresource-collections.html

Orcutt, Darby. 2009. Library Data: empowering practice and persuasion. ABC-CLIO.

Chrzastowski, Tina E., Michael Norman and Sarah Elizabeth Miller. 2009. 'SFX Statistical Reports: A Primer for Collection Assessment Librarians.'" *Collection Management*, 34(4), 286 — 303

---

[i] Cindi Trainor is the Coordinator of Library Technology and Data Services at the Eastern Kentucky University Libraries (**EKU**) and Jason Price is the Manager of Collections and Acqusitions at the Claremont Colleges Library, which serves the Claremont University Consortium (**CUC**).

[ii] If this seems far-fetched, try out the PubGet interface, http://pubget.com/ . See Chapter 4 for further discussion.

[iii] See the IOTA project, http://www.openurlquality.org for a much more extensive database level source URL quality assessment

[iv] A soapbox here: publishers like Wiley, Springer, and Elsevier that include "date published online" for each of their articles (a date that's often decades away from the actual publication date) confuse Google's automatic indexing, an confuse students to boot! We are unaware of any academic or functional reason to include tthe date an article was "published" online.

[v] See http://delicious.com/cindi/dissertations for two links that cover fixing dissertation linking in SFX.

[vi] For a more comprehensive discussion of improvements to the link resolver menu interface, see Chapter 2 and work on SFX by David Walker (http://library.calstate.edu/walker/2007/improving-the-sfx-menu/#more-26)

CHAPTER 4 – The future of OpenURL linking: Adaptation and Expansion

Abstract

Previous chapters in this report have addressed the continuing importance of OpenURL linking in libraries and presented interface-based and data-based ways to improve local OpenURL link resolver systems. This chapter explores issues pertinent to the continued and expanded adoption of OpenURL and other linking technologies, with an eye toward incorporating the shift in library collections from ownership to access and our users' desire for instant access to online full text.
=======

OpenURL link resolvers are a staple service of academic and other libraries. In 2009, ~100,000 SFX menus were presented to EKU users. Of these, ~62,000 (62%) included full text targets, ~50,000 (80%) of which were clicked. These statistics reflect the fact that EKU relies heavily on native full text on the EBSCOHost platform to satisfy the majority of user needs. In 2009 at CUC, ~283,000 of ~422,000 Serials Solutions-based searches (67%) came through the 360 Link resolver (up from 61% in 2006). The remaining requests came from the A-Z list (26%) and the OPAC (7%). These figures prove that OpenURL is the main means of library-based access to journal content at CUC. Furthermore, they represent overwhelming evidence that CUC users have shifted to dependence on the linking functionality provided by the OpenURL resolver in a relatively short period of time.

Despite this growing dependence on OpenURL, investment in its ongoing development and optimization by both vendors and libraries seems to have waned in the past few years. It is our hope that as next generation discovery tools increase the importance of OpenURL effectiveness, libraries and vendors will approach OpenURL with renewed interest and vigor. In this chapter, we present an overview of the emerging trends and technologies that may guide the ongoing development of OpenURL resolvers as they adapt to changes in the research environment and expand to serve the wider web.

Adapting to changes in the research environment

Access vs. Ownership

Online accessibility of metadata and full text content has resulted in a fundamental change in user expectations and a concomitant adjustment in library collection building principlesvi. As users discover globally-distributed content and grow to expect instantaneous access, libraries are transitioning from limited 'just-in-case' local collections to 'just-in-time' access to a wider range of content. This shift toward access over ownership has profound implications for OpenURL linking functionality.

OpenURL resolvers were originally designed to ask: does my library have this item, and, if so, how do I obtain it? The shift toward an expectation of instantaneous access and away from ownership changes this question to: how do I get this item? and how long will it take for me to do so? Resolver menus have already adapted to this change to some degree, by linking to Google and Interlibrary Loan or Document Delivery, but further change is needed to meet user expectations more effectively.

Ideally, instead of listing services through which an item can be acquired, resolver menus should indicate the delivery time for each format available. For example, instead of advising users to "Request this article via Interlibrary Loan," they should read: "Deliver this article in three days or less," as appropriate. Also, in keeping with an instantaneous, access-based approach, resolvers should be configured to support unmediated pay-per-view access to appropriate journal article collections whenever this service can be offeredvi. The ideal resolver menu for books would search locally-available catalogs and present holdings, availability and delivery information as well as Interlibrary Loan request links, where appropriate. Libraries participating in patron-driven print book acquisitions or with print-on-demand book machines could bring these options to the resolver as wellvi.

Alternate Content types

An increasing amount of research and scholarly work depends on communication in alternate formats. These range from conference proceedings and datasets, to audio and visual files, even to administrative and other non-scholarly content. The OpenURL standard is inherently flexible enough to accommodate these formats, but the resolver knowledge base is not. This suggests a general principle that should guide the future of OpenURL: use only when necessaryvi. In other words, OpenURL should only be applied to situations and content types that experience the appropriate copy problemvi. When a static link or even a specific web search will do, it is often preferable. The appropriate copy problem does not exist for content that is available only in one place (e.g. datasets) or content that is freely-available to all and therefore appropriate for all. Practical reasons, however, tend to drive the use of OpenURL for freely-available and other less apposite content; it is currently the only hook into proprietary source databases that libraries can control.

Libraries face a growing appropriate copy problem due to the wide variety of platforms that host electronic books. One simple improvement is to ensure that our e-book provider platforms are made to be OpenURL-compliant sources. Source OpenURL functionality is arguably more important for e-book platforms than for e-journal platforms, because, unlike for journal content, e-book restrictions and usability issues often drive users to want to borrow or buy a print copy. Furthermore, OpenURL is necessary to enable easy navigation from a digital-rights-restricted copy of a book (e.g. partial access, no download or limited printing) to a version with no restrictions (i.e. on the publisher's site). As of this writing, the only one of the Big Five e-book platforms used in libraries that supports OpenURL is Google Books, via its "Find in a library" link to OCLC's worldcat.orgvi.

E-books present some unique challenges for OpenURL resolver knowledge bases and vendors. Because most books are not serial publications, they have to be represented at the individual book level. Thus there are roughly two orders of magnitude more potential book records than serial records (~20 million books vs. 200,000 journal titlesvi). Although books have standard numbers assigned to them as journals do, books often have several ISBNs assigned to different physical and electronic manifestations of the same work, where journals only have one commonly-used, comprehensive identifier. OCLC's xISBN has the potential to be a major help here, but the challenge of deciding the appropriate level of distinction between intellectual works is not easy to solve. This next phase of knowledge base building is necessary though,

and differs from the first phase in that it is taking place after library integration into Google Books and Google Scholar.

Interoperability/Data exchange

The library's web page is a fractionated portal to hundreds of disparate resources that we try to get our users to take advantage ofvi. Users have had to go to different search tools to access books, e-books, journal articles, patents, etc. We present long lists for our users to navigate: lists of citation databases, individual e-journal titles, e-journal collections, primary resources, library catalogs, digital libraries, institutional repositories and more. Google's broad and deep reach has made the "library way" an increasingly harder sell. Proxy access, OpenURL, and now unified discovery tools have made some headway in addressing these issues, but we still have a long way to go.

In essence, libraries are struggling to overcome the difficulties inherent in this world of disparate online information silos. In a print-dominated world, local silos were necessary; the collections a library had on hand largely determined the universe of items available to its users. As online content and access became the norm, physical limitations on collections began to fall away, but information silos proliferated. As such, we still have to repeat the mantra: "you need to go to the library (web page) to search for this...or access that."

Web services and application programming interfaces (APIs) allow data to be pulled into catalogs and resolvers from external sources. The use of these tools reduces the need to search multiple locations as well as limiting dead ends. These tools are still constrained to the exchange of small amounts of data per transaction, and there is increasing demand for "best in class" services to provide localized, up-to-date access to the entire scope of a library's holdings. There are ExLibris customers who want to implement Serials Solutions' Summon, Serials Solutions customers who want to integrate bX; and Innovative Interfaces customers who want to present a different vendor's catalog discovery layer. These scenarios are difficult to impossible at this time, as libraries cannot successfully maintain multiple versions of their knowledge base or catalog, and vendors are slow to make their customers' data fully available and interoperable to each other. Labor-intensive workarounds to these challenges abound, but are ultimately not sustainable.

One bright spot in this landscape is the general recognition that more effective sharing of holdings data would be to everyone's advantage. Scholarly Information Strategies explored the concept of a "centralised" approach to knowledge base production in a report commissioned by UKSG in 2006vi. This model would "revolve around a single repository of content definitions and packages...that would be publicly accessible to all who desired to use it." Although such a solution would not address local customization, it would free up significant resources currently being devoted by each vendor to create the underlying knowledge base for their own products. Personal conversations with management personnel from Serials Solutions and ExLibris have confirmed that they would welcome the opportunity to redirect these resources into other means of improving their resolvers' functionality. It is our hope that the increasing demand for seamless exchange of library holdings will lead to a greater willingness to support regular exchange of knowledge base data in an interoperable formatvi.

Disaggregation of Content

Knowledge bases were designed to describe journal holdings. Journals are naturally aggregated at three levels: articles within issues within volumes within titles.   Web access is enabling disaggregation of this content: individual articles are regularly available and discoverable outside of their traditional contexts. Author web pages, institutional repositories (e.g. Harvard's DASH), open archives (e.g. PubMed Central or arXiv), and author choice open access (where authors can pay a fee to make their articles freely available within a for-fee journal), are making millions of individual articles available in a way that cannot be described at the issue level or above.

Knowledge bases, as they are currently constructed, are incapable of representing "holdings" at the article level. This limitation to knowledge base granularity has necessarily resulted in resolvers relinquishing linking of disaggregated content to web search tools like Google. As noted above, this is not necessarily a drawback, as this content does not suffer from the appropriate copy problem: its universal accessibility makes it appropriate for all. However, successful integration of search-engine access to this content into resolver menus is an ongoing challenge (see Chapter 3).

Complementary systems

When referring to OpenURL's direct "competition" in library instruction sessions, one of the authors of this report often refers to the ever-growing number of static links as "fast and dumb" and OpenURL links as "slow and smart". PubMed and Google Scholar, for example, have links that go directly to publisher content, whether it is licensed or not. These static links are preferable when the content is available, but are a dead end when the content is not. These links are necessary for independent users and users whose library does not have an effective resolver.

Since static links inherently point to a single location, OpenURL links are necessary to provide users access to non-publisher-direct content from aggregators or on the open web. As such, libraries should seek to make their resolver complement these static links by ensuring that they are offered alongside them, and that the resolver links work as effectively as possible. In the same vein, resolvers should be altered to include static links whenever they are the most appropriate (or only way) to access the content. This perspective, then, reflects a common theme of this report: resolvers must provide access to as broad a range of content as possible as accurately as possible, lest our users lose faith in their utility.

DOI, the Digital Object Identifier, was developed about the same time as the OpenURL. DOI linking depends on a linking service called CrossRef, which is a registration agency of the International DOI Foundation. DOIs are a way to assign persistent unique identifiers to online objects and can be one piece of metadata transported in an OpenURL. In a sense, a DOI is a hybrid between a static link and a knowledge base-driven OpenURL link. They improve on static links because they are stable persistent identifiers. They are similar to OpenURLs in that they depend on a directory of content. The DOI directory contains the DOI, citation metadata, and item URL. Publishers can update the item URL at any time when the address of the object changes. It is important to note, however, that CrossRef does not maintain library knowledge base datavi.

Libraries use the DOI/CrossRef system in two main ways: to retrieve DOIs that are integrated into their resolver menus, thus providing a direct link to publisher's full text; and to retrieve the bibliographic metadata for a known DOIvi. Unfortunately, the implementation of the first case, as tested by the authors, leaves much to be desired. CrossRef links to publisher full text failed 25% of the time and were redundant in nearly every other case (see Chapter 3). However, an extension of the second case is of crucial importance in a way not previously recognized by the authors. CrossRef has provided a means whereby DOIs on the web can serve as source URLs, enabling OpenURL linking from hundreds of publishers' content. We describe this functionality in detail below. It is important to emphasize that CrossRef/DOI functionality is a complement rather than an alternative to OpenURL. It cannot address the appropriate copy problem without referring to the library knowledge base by means of an OpenURL resolvervi.

Seamless connectivity

One vision of the ideal future of OpenURL link resolution involves its continued progression from foreground service to background functionality. It should, perhaps, be our goal to render as *few* resolver menus as possible, replacing them with one-click direct linking to the best full text version available. As discussed in Chapter 3, this functionality is currently available from both major resolver vendors, although its breadth and reliability need improvement.

Another ideal complement to one-click delivery of full text would be indication of full text availability via the resolver button in the source database. There are two levels of possible functionality here. First, as the button is being rendered, the source database could query the resolver knowledge base for full text availability and insert a "get full text" version of the button whenever it finds a match, instead of the standard resolver button. Similar functionality is built in to the ExLibris MetaLib results set; this highly desirable feature should be implemented for other sources, wherever possible. The authors hope that a future iteration of link resolver software or its successor will confirm full text access before providing links to the user.

This vision and functionality have been realized in pubget (http://pubget.com), the first implementation of an OpenURL-based 'pull' technology in a search tool. Back in the early days of OpenURL it was magical just to be able to follow the path from result/citation to full text (in any number of steps) without having to manually translate citation metadata. The next generation of OpenURL integration may obviate the need to follow a path at all, inserting full text into the search process, rather than requiring users to leave the search interface to hunt for full text (which may or may not be available to them).

Rather than pushing the user out to the full text via a bewildering (or at least distracting) plethora of paths, pubget pulls in PDFs and co-locates them with the search result list. At first blush, their website seems to provide a magical service, free to not-for-profit organizationsvi, complete with all the secrets that make magic what it is. Behind the scenes, it is a knowledge base- and resolver-driven service that reduces the number of steps from discovery to delivery to zero (when the PDF is available).

Of course, pubget has its limitations. The universe of 25 million citations it searches consists only of PubMed, ArXiv, and JSTOR records. Like any link resolver, pubget's accuracy is limited by the quality of the knowledge base on which it's based. Some pubget libraries knowledge bases were found to have accuracy levels as low as 70% (M. Abrams, pers. comm.) The company is actively developing strategies to increase library-level knowledge base accuracy by augmenting their version with library-specific, direct-from-publisher access lists. As of March 2010, pubget chose to stop accepting new customers, instead focusing on the accuracy of PDF retrieval for its current 220 libraries. The impact of pubget for libraries is still uncertain, but it gives us a glimpse of a future where a link resolver functions fully behind the scenes.

Expanding the reach of reference linking: OpenURL on the web

An increasing amount of research starts with web search enginesvi. Even research that starts at a library website or citation database quickly gets funneled away, because such a high percentage of content is hosted beyond the libraries domain. As users conduct more research on the open web, it has become crucial for libraries to ensure that users have access to the

high-quality, library-funded content from the place where they spend the majority of their research time.

OpenURL resolver functionality has yet to establish a significant presence outside of proprietary library indexes. Google Scholar, PubMed, Google Books and Open WorldCat are the major exceptions to this blanket statement, yet compared to the web as a whole, even these behemoths are quite small. The most significant challenge in the future of OpenURL is expansion onto the Web. The range of this expansion must include both the bibliographies of full text items contained in library-funded collections *and* citations and bibliographies available on the open web.

The technological infrastructure necessary to support an expanded reach of OpenURL already exists; its greatest challenge is adoption and implementation. Two requirements must be met to enable OpenURL linking from citations on the web. The citations must be coded with OpenURL-compliant tags or DOIs, and web browsers must be extended to identify these codes and insert an affiliation-aware resolver button. The following three sections describe existing technology that supports these requirements, and offer specific suggestions for meeting them.

Enabling OpenURL linking from DOIs on the web

CrossRef has registered more than 40 million metadata records for scholarly itemsvi. Many of these items are cited in multiple places on the web. Libraries can facilitate access to this content from any web index or bibliography that includes DOIs.

The default behavior of DOI links on the web is to direct users to the publisher's full text. In many cases, users will not be authenticated for this access, either because they are working outside of their library's network, or because their library does not license access to the publisher version of the item. Libraries have the option to configure the CrossRef server to send DOI requests through their library's link resolver rather than directly to the publisher full textvi. To accomplish this, a library registers its resolver base URL with CrossRef. Once they do so, a persistent cookie is downloaded that contains the URL for the local resolver server.

This cookie enables OpenURL for DOIs within the browser, which will lead the CrossRef system to redirect DOI requests to the local resolver. The local link resolver then receives the metadata needed for link resolution, either from the source of the link or from the CrossRef DOI directory. Unfortunately, neither of the authors can vouch for the effectiveness of this service, as we have yet to implement it at either of our institutions, although we can test it through LibX-enabled right-click context menus (see *Leveraging COinS*... below). Since this configuration replaces direct linking with resolver-based linking, it will be important for libraries to confirm that activating it will increase full text access for users. Ultimately, the extensive reach of this service into the bibliographies of millions of articles on the web will justify its implementation.

COinS to enable the web where DOIs aren't present or available

COinS is an acronym meaning "Context Object in SPAN" and is a way for web content creators to embed citation information into any web page using an HTML <SPAN> element. Users must install software such as LibX or OpenURL Referrer to make a browser COinS-aware. When the browser is operating from within an IP range or proxy server IP that is registered with OCLCs WorldCat Registry, it will automatically be directed to the libraries local link resolver. When a COinS-aware browser encounters a COinS <SPAN> element, it places a resolver button in place of the code. Thus, COinS is a way to create OpenURLs that are tied on the fly to a specific resolver each time an HTML page containing COinS code is served. With COinS, a resolver button can appear anywhere there is coded citation data. COinS is currently utilized by reference managers (including Refworks, Zotero and Mendeley), by a few publishers, and is embedded in HubMed (sic), WorldCat records and in many Wikipedia pages.

COinS code looks like this:

```
<span class="Z3988" title="ctx_ver=Z39.88-
2004&amp;rft_val_fmt=info%3Aofi%2Ffmt%3Akev%3Amtx%3Ajournal&amp;rfr_id=info%3Asi
d%2Focoins.info%3Agenerator&amp;rft.genre=article&amp;rft.atitle=Linking+and+the+Ope
nURL&amp;rft.title=Library+Technology+Reports&amp;rft.stitle=LTR&amp;rft.issn=0024-
2586&amp;rft.date=2006&amp;rft.volume=24&amp;rft.issue=1&amp;rft.spage=1&amp;rft
```

.aulast=Grogg&rft.aufirst=Jill&rft.au=Jill+Grogg&rft_id=info:doi/10.1108%2
F00330330910978581">

    (descriptive text, or remove so that only the resolver button displays)

It is easily generated and embedded into any library webpage_vi. This is useful for institutional repositories, faculty profile pages and learning management systems, as well as for library blogs, wikis and new book lists. COinS support is also being built into open source systems used in various libraries such as Drupal modules, the open source next-generation catalog software Scriblio, and the popular blogging platform WordPress. We strongly encourage libraries to invest effort in providing services to their faculty by embedding COinS code in strategic places. COinS coding of publications listed on faculty profile pages will make it easier for researchers and prospective students to find a copy of the item that is available to them. COinS coding of items deposited in institutional repositories will facilitate access to an authoritative copy of manuscripts and pre-prints.

Leveraging COinS coding: key browser extensions

The COinS extension with the most impact on library researchers is a browser extension called LibX. Developed at Virginia Tech by Annette Bailey and Godmar Beck, LibX comprises several parts that together make for a powerful research experience. In addition to COinS support, LibX facilitates searching the library catalog, electronic journal list and other resources from a toolbar or from the right-click context menu; ISSNs and ISBNs found on any web page are linked to a library catalog search; any web page can be reloaded through the library's proxy server; there is support for drag-and-drop Google Scholar searching; and visual cues linked to the library catalog are embedded in Amazon.com and other websites. LibX is currently available for Firefox, Internet Explorer and Chrome, and requires local installation.

OpenURL Referrer is a much simpler extension which is also available for Firefox and Internet Explorer, but is not compatible with the latest version of Firefox (3.6.8) at the time of this writing. Furthermore its resolver functionality is less favorable than LibX in that the resolver buttons are not locally-=branded and require more clicks to get to full text.

COinS is also utilized by reference management software--Endnote, RefWorks, Mendeley and Zotero, to name a few--making it easy for a researcher to return to the full text of any item as provided to him or her by the library from their collection of references. These programs support download of tagged citations via an icon in the address field or toolbar and/or via bookmarklets. Bookmarklets can extract citation metadata from COinS- or DOI-coded pages, and will even create a less structured web page citation for pages with scant metadata. Pubget has extended bookmarklet functionality to direct PDF retrieval, allowing users that don't use reference management software to retrieve PDFs from abstracts in PubMedvi.

Other linking initiatives

As has always been the case with the web, OpenURL is not the only linking technology, but it does solve a particular problem, that of connecting and uncovering sometimes-hidden library holdings. Other linking initiatives that may influence the future of article and other item linking in the library landscape include the semantic web and microformats.

The Semantic Web The web as we know it today consists of links that work and break instantaneously and that carry no indication of the relationship between one object and another. Information on the web today is still largely text-based rather than based in machine-readable data. Simply put, humans can derive meaning from the words on a web page, but computers cannot. The phrase Semantic Web encompasses efforts to create a framework for bringing machine-readable meaning (semantics) to the web.

Efforts to bring bibliographic data into the Semantic Web are described succinctly and accessibly by Karen Coyle in her two *LTR* issues, "Understanding the Semantic Web: Bibliographic Data and Metadata" and "RDA Vocabularies for a Twenty-First-Century Data Environment." While it is easy to envision the application of the Resource Description and Access cataloging rules in physical libraries as they exist in the early 21st Century, it is less clear how RDA might extend to apply to and help retrieve items not necessarily collected individually in a library, yet available and desired by our online users: articles, book chapters, dissertations, proceedings, data sets, audio and video. Regardless of this lack of clear path, putting our bibliographic data in a machine readable framework that is more

"data-like" than the current, text-heavy MARC format is a step toward making that data available for use by non-library entities on the webvi.

Microformats constitute one effort to add structure and machine-readable context to information contained in web pages. At this time, software must be added to the web browser so that microformats can be seen. The Operator plug-in for Firefox creates a toolbar that pulls out Contact, License, Event and other microformat data and can export or send it as a search to other websites. There are also extensions for Chrome, Safari and Internet Explorer, though the Chrome extension detects and displays only the hCard microformat at the time of this writing.

A draft specification of the Citation Microformat exists. It is similar to COinS in that a microformat can easily be embedded in any html page for others to use. Karen Coombs writes that COinS and the citation microformat differ in that the latter will "break the data down into component parts to make it more flexible" rather than building on the OpenURL Context Objectvi. As the microformats.org citation formats page shows, there are myriad ways to display citation metadata; the discussion to create a single hCitation microformat is likely to be long and complicatedvi.

Conclusion

It has been interesting to watch the migration of library content to the web and the evolution of the tools that libraries devise and purchase to connect their users with that content. Users, meanwhile, have turned in droves to Google and other free web tools for their research needs. Rather than making libraries irrelevant, users' attraction to tools like Google have challenged us to make quality information available conveniently, quickly, and simply.

CHAPTER 5 - SOURCES & RESOURCES

OpenURL

Beit-Arie, Oren, Priscilla Caplan, Miriam Blake, Dale Flecker, Tim Ingoldsby, Laurence Lannom, William Mischo, Edward Pentz, Sally Rogers, and Herbert Van de Sompel. "Linking to the Appropriate Copy: Report of a DOI-Based Prototype." D-Lib Magazine, 7:9. 2001. [http://www.dlib.org/dlib/september01/caplan/09caplan.html] Accessed 12/3/2003; URL verified 7/28/10.

Chandler, Adam. "Results of L'Annee Philologique Online OpenURL Quality Investigation." http://bit.ly/chandler-mellon

Grogg, J. E. (2006). "Linking and the OpenURL." *Library technology reports*, 42, 2006. http://bit.ly/LTRgrogg

> This report serves as an excellent introduction to the need for and development of OpenURL, as well as a summary of products for and uses of linking available in 2006.

KBART: Knowledge Bases and Related Tools Working Group. http://www.niso.org/workrooms/kbart. Accessed July 28, 2010.

KBART Phase I Best Practices. http://www.niso.org/publications/rp. Accessed July 30, 2010.

KBART 5.3.2.1: Data Fields and Labels. http://www.uksg.org/kbart/s5/guidelines/data_field_labels. Accessed July 28, 2010.

KBART Registry. http://sites.google.com/site/kbartregistry/ Accessed July 28, 2010.

LeBlanc, Jim. "Measuring the Quality of OpenURLs: An Interview with Adam Chandler." *Information Standards Quarterly*, 22, 2010. http://bit.ly/leblanc-chandler Accessed July 28, 2010

IOTA, Improving OpenURLs Through Analytics. National Information Standards Organization. http://www.niso.org/workrooms/openurlquality (accessed 7/30/2010)

The Semantic Web & COinS

Bailey, Annette and Godmar Beck. 2008. "Retrieving Known Items with LibX." The Serials Librarian. 53(4), p 125-140

COinS ocoins.info is the official website for COinS information, including specifications, implementation guidelines, links to software and sites that use COins, and a COinS generator useful for creating code to embed on any web page. (accessed 7/30/2010)

Chudnov, Daniel. "COinS for the Link Trail." *Library Journal netconnect*, summer 2006, page 8.

Coombs, Karen. "Microformats: Context Inline." *Library Journal*, April 15, 2009. p64

Coyle, Karen. "Understanding the Semantic Web: Bibliographic Data and Metadata." *Library Technology Reports* January 2010.

Microformats specifications: http://microformats.org (accessed 7/30/2010)

Browser Extensions

LibX: http://libx.org (accessed 7/30/2010)
OpenURL Referrer: http://www.openly.com/openurlref/ (accessed 7/30/2010)
Operator (Firefox): http://microformats.org/wiki/Operator (accessed 7/30/2010)
Chrome Microformats extension:
https://chrome.google.com/extensions/detail/igipijakdobkinkdmiiadhghmbjhciol

SFX
Brooks-Keiffer, Jamene. "Working the Workaround: DTFT Local."
http://ksulib.typepad.com/sfxdoc/2010/01/working-the-workaround.html Accessed July 28, 2010.

Chrzastowski, Tina E., Michael Norman and Sarah Elizabeth Miller. 2009. 'SFX Statistical Reports: A Primer for Collection Assessment Librarians.'" *Collection Management*, 34(4), 286 — 303

Cummings, Joel and Ryan Johnson. "The Use and Usability of SFX: Context-Sensitive Reference Linking." *Library Hi Tech*, 21:1. 2003.

Stowers, Eva and Cory Tucker. "Using Link Resolver Reports for Collection Management." *Serials Review* v 35 no 1. 2009.

Jina Choi Wakimoto, David S. Walker, Katherine S. Dabbour, The Myths and Realities of SFX in Academic Libraries, *The Journal of Academic Librarianship*, Volume 32, Issue 2, March 2006,

Pages 127-136, ISSN 0099-1333, DOI: 10.1016/j.acalib.2005.12.008.
(http://bit.ly/JALwakimoto)

Walker, David. "Chameleon SFX Catalog Integration Plugin."
http://www.exlibrisgroup.org/display/SFXCC/Chameleon+SFX+Catalog+Integration+Plugin
(Accessed 7/30/10)

Walker, David. "Improving the SFX Menu."
http://library.calstate.edu/walker/2007/improving-the-sfx-menu/#more-26 (Accessed 7/30/10)

Walker, David. "Integrating Print Holdings into SFX." *Journal of Interlibrary Loan, and Electronic Reserve*. 15(3) p95-108.

Usability and User Experience

Alertbox from Jacob Neilsen: http://www.useit.com/alertbox/

Adaptive Path blog. http://www.adaptivepath.com/blog/

King, David Lee. 2008. *Designing the Digital Experience: How to Use Experience Design Tools and Techniques to Buid Websites Customers Love*. Medford, New Jersey: Information Today.

Krug, Steve. *Don't Make Me Think*. New Riders Press; 2nd edition. 1995. Also see the companion website, http://www.sensible.com/dmmt.html for a sample usability script and other resources.