

Digging into the Data

Exposing the Causes of Resolver Failure

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.¹ 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 at our libraries. The majority of these tasks are broadly applicable, and many can be applied individually to improve resolver effectiveness at any library.

Testing OpenURL Full Text Link Resolution Accuracy at Our Institutions

This study is based on the “real-life” approach of Wakimoto and others (2006) to allow a historical comparison with their 2004 SFX testing results.² 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 3). Citations that included native full text were avoided, as well as those from journals or books that had been tested previously.

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 4). 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, Walker, and Dabbour's designations.³ Their results are included for comparison (table 5).

Wakimoto and others (2006) reported that about 20 percent 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. For this study, however, the category “Required search or browse for full text” was reassigned from the Correct group to the Error group to reflect reduced user willingness or ability to further navigate to 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 percent, averaged across both datasets. This results in total error rates of 35 percent for the Wakimoto and others dataset and 29 percent for our dataset (table 5).

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 percent)

were available via the Web. Tapping into this content is equivalent to increasing our budgets by 15 percent. 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 webpages and in institutional repositories. This additional category of false negatives increases the overall error rate to 33 percent. 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 document⁴) 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 percent of the time (in 86 of 93 menus), and the ECU resolver was successful 70 percent of the time (in 54 of 77 menus). Thus, by this measure, the resolvers were 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 percent 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 6). Unfortunately, the vast majority of these successes (101 of 105) reported negative

Database	Document Type*					Total
	BC	BK	JA	NA	DT	
AH&L (Ebsco)	10		20		10	40
ASP (Ebsco)			20	10		30
Eric (CSA)		10	20		5	35
MLA (Ebsco)**	5	5	10		5	25
MLA (CSA)**	5	5	10		5	25
NCJRSA (CSA)**	5	5	10		5	25
PsycInfo (Ebsco)	10	10	20		10	50
Scholar (Google)			20			20
SocAbs (CSA)	10	10	20		10	50
Summon (SerSol)**	5	5	10	5	5	30
Worldcat.org (OCLC)**		5	10		5	20
Total	50	55	170	15	60	350

Table 3
Number of citations (source OpenURLs) tested by database and document type

Institution	Source URLs tested	# of menus w/o FT links	% of menus w/o FT links	# of menus w/ FT links	% of menus w/ FT links	# of FT links tested	Average# of FT links	% of menus w/ >1 FT link
CUC	166	74	45%	92	55%	212	2.30	75%
EKU	185	108	58%	77	42%	131	1.70	31%
Total	351	182	52%	169	48%	343	2.03	55%

Table 4
Number and proportion of menus with full text links offered by each institution.

* BC = book chapter, BK = book, JA = journal article, NA= newspaper article, DT = dissertation
 ** These database/vendor combinations were tested at only one of the two libraries in this study.

results, reflecting small e-book collections or their absence from the knowledge base. In addition, because the study was designed to emphasize book content (40 percent 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 percent (270 of 420 results). With this in mind, our results show that only about two out of three non-book resolver results are accurate.

In contrast to book content, newspaper and dissertation results had much lower accuracy rates than average (0.38 and 0.30, respectively, table 6). 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 were 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 causes of failure, below.

Category	Dataset							
	2004 CSU Northridge & San Marcos		2010 CUC & EKU		2010 CUC		2010 EKU	
Correct—No FT available	94	36%	138	26%	57	20%	81	34%
Correct—sent directly to FT	45	17%	86	16%	49	17%	37	15%
Correct—sent to citation with FT Link	29	11%	147	28%	112	39%	35	15%
Total Correct	168	65%	371	71%	218	76%	153	64%
Error—Required search or browse for FT	39	15%	59	11%	30	10%	29	12%
Error—menu says we have it, but don't	29	11%	51	10%	21	7%	30	13%
Error—menu says we don't have it, but do	24	9%	44	8%	17	6%	27	11%
Total Error	92	35%	154	29%	68	24%	86	36%
Total	260		525		286		239	

Table 5
Resolver results for full text requests in each dataset (after Wakimoto and others, 2006).

Database	Document Type*										Totals
	BC		BK		JA		NA		DT		
	Succ.	Fail	Succ.	Fail	Succ.	Fail	Succ.	Fail	Succ.	Fail	
America: History & Life	10				28	1			1	9	49
Academic Search Premier					20	11	11	26			68
ERIC			7	3	27	5			3	2	47
MLA (Ebsco)	5		5		9	5			1	4	29
MLA (CSA)	5		5		23	4			2	3	42
Nat'l Crim Just Ref Svc Abs	4	1	5		2	9			1	4	26
PsycInfo	10		10		23	5			2	8	58
Google Scholar					36	17					53
Sociological Abstracts	10		10		31	4			1	9	65
Summon	5		5		20	7	7	3	3	2	52
Worldcat.org			5		15	11			4	1	36
Totals	49	1	52	3	234	79	18	29	18	42	525
Accuracy Rate	0.98		0.95		0.75		0.38		0.30		
Overall Accuracy Rate	0.71										

Table 6
Resolver full text link accuracy rate by document type and source database. An interactive version of this table that allows examination of the details of the specific results represented by each cell is available online (<http://bit.ly/openurltables2010>).

* BC = book chapter, BK = book, JA = journal article, NA = newspaper article, DT = dissertation

Nearly two thirds of the results were for journal articles, so perhaps not surprisingly, their accuracy rate most closely mirrored the overall results (75 percent, table 6). 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 journal articles could be tested across all citation databases, and half of the database/vendor combinations were tested at only one institution. Finally, the IOTA project (Improving OpenURLs through Analytics) 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.

IOTA (Improving OpenURLs through Analytics)
www.openurlquality.org

Causes of Failure

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 obscures a 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.

Failure Cause Analysis Procedure

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.

Cause of Failure	Error Type				Total	%	Total -DT	% -DT
	False Pos.	False Neg.	Req'd search or browse	Cause %				
Source URL data inaccurate	10	10	6	0.17	33	0.22	33	0.26
Source URL data incomplete			7	0.05				
Resolver KB inaccuracy	16	4		0.13	51	0.33	23	0.18
Resolver translation error		28	3	0.20				
Resolver target URL incomplete / Provider doesn't accept item level links			21	0.14	21	0.14	21	0.17
Provider target URL translation error			18	0.12	33	0.22	33	0.26
Provider content incomplete	15			0.10				
Miscellaneous	10	2	3	0.10	15	0.10	15	0.12
Total	51	44	58		153		125	

Table 7

Frequency of failure causes by error type. An interactive version of this table which allows examination of the details of the specific results represented by each cell is available online (<http://bit.ly/openurltables2010>).

In order to identify the cause of each resolver failure, a wide range of data was collected for each full text resolver result. 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 at 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.⁵

Failure Causes by Error Type

In general, the causes of resolver failure were evenly distributed across the OpenURL resolver chain. No more than 20 percent fell into any of the eight categories (table 7, column 5), and no more than 33 percent were due to any of the five components (table 7, column 7). In fact, when the 28 resolver translation errors that were due to dissertation citations were dropped from the analysis, no single component was responsible for more than 26 percent of the errors (table 7, column 9). Despite this even distribution of causes, some interesting general patterns emerge, particularly when the causes are analyzed by vendor/database and document type.

It is important to note here, however, that there are two cause categories that could not be assigned to one of the three resolution constituents (i.e., data source,

resolver, 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 versus 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 (case 1) or by replacing the journal-level link with an item-level link to search Google Scholar (case 2).

Cause of Failure	Vendor/Database													Total
	CSA				CSA Total	Ebsco				Ebsco Total	Google Scholar	OCLC World cat	SerSol Summon	
	ERIC	MLA	NCJRS	SocAbs		AH&L	ASP	MLA	Psyc Info					
Source URL data inaccurate			4	5	9						9	8		26
Source URL data incomplete			7		7									7
Resolver KB inaccuracy	1		2		3		9	1	2	12	2		3	20
Resolver translation error		3		4	7	9	2	4	8	23			1	31
Resolver target incomplete/ Host doesn't accept item level links	2		1	1	4	1	8	1	1	11	2	3	1	21
Host target URL translation error		2			2		9			9			7	18
Host content incomplete	5	1			6		7	2		9				15
Miscellaneous	2	1		3	6		1	1	2	4	4	1		15
Total	10	7	14	13	44	10	36	9	13	68	17	12	12	153

Table 8

Frequency of failure causes by source vendor and database. An interactive version of this table which allows examination of the details of the specific results represented by each cell is available online (<http://bit.ly/openurltables2010>).

Failure Causes 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 had spotless records, while CSA, Google, and OCLC produced all the errors (table 8). 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 content was also free of errors, despite the dual institution sample for three of the four 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.⁶ It is not surprising that Google Scholar had a number of source URL errors, given its crawler-based indexing approach.⁷ The high ratio of source errors from the results tested from OCLC Worldcat.org (from a single institution) may reflect lower quality indexing in 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 (70 percent) of the 33 errors 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 Causes by Document Type

The last level of failure cause analysis examines the 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.

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 9). 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 Ex Libris's 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

Cause of Failure	Document Type*					Total
	BC	BK	JA	NA	DT	
Source URL data inaccurate			16		10	26
Source URL data incomplete			7			7
Resolver KB inaccuracy	1		10	7	2	20
Resolver translation error			3		28	31
Resolver target incomplete / Host doesn't accept item level links			14	7		21
Host target URL translation error			6	12		18
Host content incomplete		3	8	2	2	15
Miscellaneous			15			15
Total	1	3	79	28	42	153

Table 9 Frequency of failure cause by document type. An interactive version of this table which allows examination of the details of the specific results represented by each cell is available online (<http://bit.ly/openuritable2010>).

* BC = book chapter, BK = book, JA = journal article, NA = newspaper article, DT = dissertation

Abstracts (5 of the 10), these can be resolved by matching the publisher data in the source URL (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, the library can to customize source URL resolution by editing the source parser.⁸ 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 9). 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 approximately 20 percent of the errors encountered (28 of 153), they appear to be quite common, since they resulted from only 4 percent 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 9). Furthermore, they were quite evenly distributed: at least 16 percent were attributed to each of the five 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



Figure 5
EKU: “This item is not available online.”



Figure 6
CUC: Long resolver menu with link to search Google Scholar placed near the bottom.

the data presented above. As active users of the product, we noticed a number of aspects of the front-end functionality

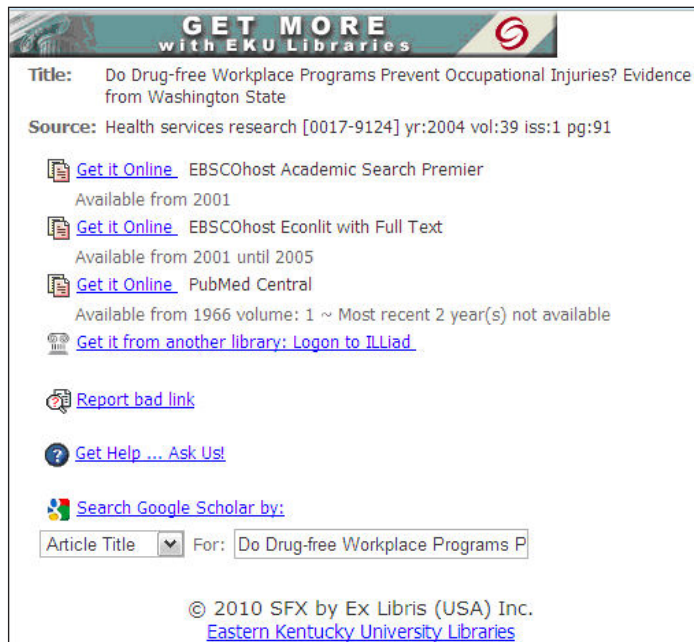


Figure 7
EKU: Duplicate links to the same article on EBSCOhost.



Figure 8
Links clicked from WorldCat.org retain a banner that enables users to return to WorldCat. The banner also includes citation information.

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 2. We present them there 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, “This item is not available online” (figure 5). 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, users generally do 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 colocated 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 (figure 6).

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 (figure 7) 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 may not affect CrossRef links.

Order of link presentation is a thornier issue. It would improve the user experience to be able to order links by some combination of **link reliability**; **link depth**; e.g., article-level versus journal-level; and **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).

- **Link 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 (see chapter 4) may have unique insight into these numbers.

- **Link depth** should be consistent within a particular provider, so it would be 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 (as WorldCat does, see figure 8).
- 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 section on Industry Initiatives in chapter 1). 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.

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 print and online availability in the resolver menu. Furthermore, both menus are set up to search the local and union (Inn-Reach) catalog in separate steps (figure 9), 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 (figure 10) is preferable, although sending searches as phrases, (i.e., in quotation marks) would improve their

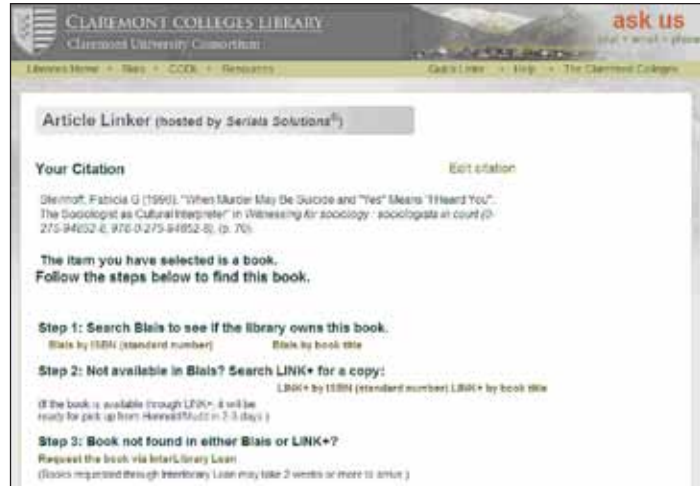


Figure 9
CUC: Menus are set up to search the local and INN-Reach union catalog in separate steps.

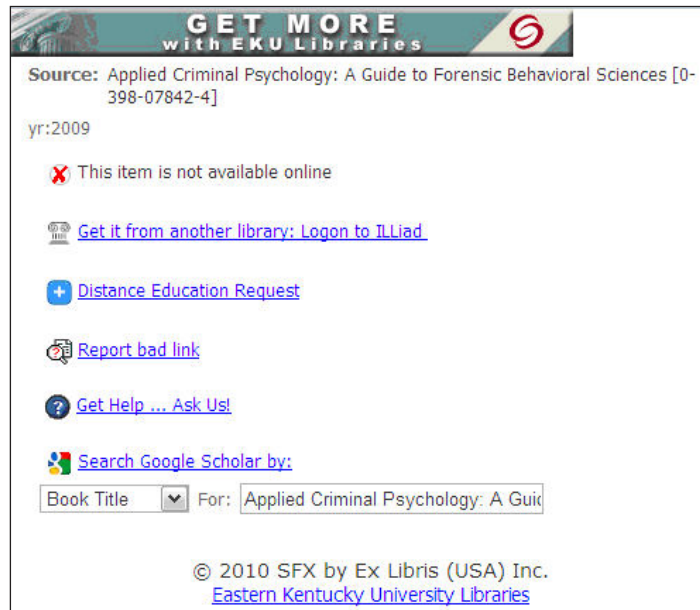


Figure 10
EKU: Menu configured to enable a Google Scholar search by book title.

results. Google offers results for keywords when the phrase search produces no results, so nothing is lost by sending the search in this manner.⁹

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.

1. **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 5), 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.
2. **Fix dissertation target linking.** EKU’s usage and OpenURL failure data provide powerful justification to fix linking to this class of resource (see tables 2 and 6). 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 Ex Libris. In the meantime, locally hosted SFX implementations can edit their source parsers to fix this problem.¹⁰ Our results showed that newspaper article linking failed almost as often as dissertation linking. Although newspapers are at least as significant a concern, their pagination and date variation, short nondistinct article titles, and frequent supplementary sections make them much more of a challenge.
3. **Review every full text provider for item-versus 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-versus title-level linking and ensuring that successful item-level linking is established wherever possible. Furthermore, knowledge of this attribute is essential for establishing the order in which full text links are presented.
4. **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 link reliability, link depth, and

format(s) available (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 that should receive the same rank and minor adjustments require extensive reranking. Perhaps a simple solution would be for Serials Solutions to change its system to allow priorities (i.e., 1, 2, or 3) rather than a ranking (1 to 314 for CUC), or even to offer its own order based on the factors above. SFX is significantly simpler to configure: it requires only 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 (e.g., “O.K.” and “bad”).

5. **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 maintained.¹¹ 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).¹² 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 “Qualitative Observations on Resolver Effectiveness” 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 in representing e-books in their catalogs, the best near-term solution is probably an adaptation of David Walker’s Chameleon SFX plugin to integrate e-book lookup into the full text services section. A similar JavaScript-based tool could potentially be built for 360 Link.

Chameleon SFX Catalog Integration Plugin

www.exlibrisgroup.org/display/SFXCC/Chameleon+SFX+Catalog+Integration+Plugin

6. **Optimize top 100 most requested journals.**

According to the 80/20 rule, 80 percent of use occurs in 20 percent of the titles, so focusing on heavily used journals will address a great deal of the overall usage. Although only SFX provides a report that 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 is collected in a systematic way, spreadsheet pivot tables can be used both to examine frequencies and to show details from individual categories.¹³ 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.

7. **Optimize top ten full text target providers.** The number of click-throughs per target host (table 1, SFX Query 7) can be approximated with Serials Solutions’ “Click-Through Statistics by Title and Database (Holdings)” report.

8. **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 article by Chrzastowski and others).¹⁴ 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=” provides 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 (such as Funnel Web) can extract elements and reveal source platform and provider publisher 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.¹⁵

Funnel Web

www.quest.com/funnel-web-analyzer

9. **Optimize top ten source databases by content type.** Once staff at a library extract 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 be discovered in standard usage reports.

10. **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 (e.g., see discussion of Pubget in chapter 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 step. 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

1. 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 Acquisitions at the Claremont Colleges Library, which serves the Claremont University Consortium (CUC).

2. Jina Choi Wakimoto, David S. Walker, and Katherine S. Dabbour, "The Myths and Realities of SFX in Academic Libraries," *The Journal of Academic Librarianship* 32, no. 2 (March 2006): 127-136.
3. Ibid.
4. If this seems far-fetched, try out the Pubget interface, <http://pubget.com>. See chapter 4 for further discussion.
5. For an exhaustive representation of the data we collected, see the AllData worksheet in the MS Excel workbook available from <http://bit.ly/openurltables2010>.
6. See the IOTA project, www.openurlquality.org, for a much more extensive database-level source URL quality assessment.
7. We'd like to climb on our 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, and confuse users as well. We are unaware of any academic or functional reason to include the date an article was "published" online.
8. For information on fixing dissertation linking in SFX, see Geoff Sinclair, "SFX and Dissertations," updated Oct. 28, 2009, Spotdocs website, <http://spotdocs.scholarsportal.info/display/sfxdocs/SFX+and+Dissertations> (accessed Aug. 4, 2010), and Jamene Brooks-Kieffer, "Working the Workaround: DTFT Local," Jan. 21, 2010, K-State Libraries website, <http://ksulib.typepad.com/sfxdoc/2010/01/working-the-workaround.html> (accessed July 28, 2010).
9. For a more comprehensive discussion of improvements to the link resolver menu interface, see chapter 2 and work on SFX by David Walker ("Improving the SFX Menu," Jan. 3, 2007, <http://library.calstate.edu/walker/2007/improving-the-sfx-menu/#more-26> [accessed July 30, 2010]).
10. Brooks-Kieffer, "Working the Workaround."
11. Chad Hutchens, "Managing Free and Open Access Electronic Resources," *UKSG Serials-eNews*, no. 210 (Dec. 11, 2009), www.ringgold.com/UKSG/si_pd.cfm?AC=0350&Pid=10&Zid=5067&issueno=210 (accessed Aug. 4, 2010).
12. Jamene Brooks-Kieffer, "ER&L 2009: Managing Free E-resource Collections," Feb. 12, 2009, K-State Libraries website, <http://ksulib.typepad.com/conferences/2009/02/erl-2009-managing-free-eresource-collections.html> (accessed Aug. 4, 2010).
13. See the tables in the MS Excel workbook available from <http://bit.ly/openurltables2010>.
14. Tina E. Chrzastowski, Michael Norman, and Sarah Elizabeth Miller, "SFX Statistical Reports: A Primer for Collection Assessment Librarians," *Collection Management* 34, no. 4 (2009): 286-303.
15. See, for example, Darby Orcutt, *Library Data: Empowering Practice and Persuasion* (Santa Barbara, CA: ABC-CLIO, 2009).