

## **NISO IOTA: Improving OpenURLs Through Analytics, in Context**

by Adam Chandler

Database Management & E-Resources Research Librarian  
Central Library Operations  
Cornell University Library  
110B Olin Library  
Ithaca, NY 14853-5301  
607-255-5760  
[alc28@cornell.edu](mailto:alc28@cornell.edu)  
@alc28 on Twitter

Submitted for publication in *Against the Grain*, December 2010.

The OpenURL 1.0 specification was finalized in 2004 (National Information Standards Organization, 2004). The research that underpins OpenURL reaches back into the 1990s, when Herbert Van De Sompel, working with colleagues at Ghent University, demonstrated an alternative to static, bidirectional linking: dynamic reference linking (Van de Sompel and Hochstenbach, 1999). The problem Van De Sompel and others had to solve was how to break out of the fragile, proprietary bi-lateral linking relationships between licensed content providers to make linking to the "appropriate copy" (Beit-Arie et al. 2001.) possible. OpenURL was a brilliant and elegant solution to the appropriate copy problem. Format a URL using standard name/value pairs and send it to a library link resolver. That is, shift the burden of maintenance. Have the link resolver -- software designed for the task -- figure out how to link to the full text content, based on the local library holdings. The solution looks obvious to us in 2011 because OpenURL linking is so pervasive. Thousands of libraries around the world use link resolvers. By one estimate, some 1 billion OpenURLs are sent to link resolvers annually (HangingTogether Blog, 2009). It is an integral part of the library technology fabric, on par with an OPAC or A-Z ejournal list. There is a problem, however. OpenURL links fail. Frequently. In 2007 UKSG commissioned a survey that exposed the extent of this problem. James Culling noted:

"72% of respondents to the online survey either agreed or strongly agreed that a significant problem for link resolvers is the generation of incomplete or inaccurate OpenURLs by databases (for example, A&I products). OpenURLs may be broken on account of insufficient or incorrect metadata that leads to erroneous results in the link

resolver's service menu or prevents the resolver from creating a sufficiently deep link to a target site. One librarian interviewed commented that his experience with some sources was so bad that he refused to enable OpenURL links from them as he did not wish to expose his end users to the problems" (Culling, 2007, p. 33).

Recently, Trainor and Price (2010) dissected the errors they observed in a random sample of OpenURL requests. Their careful testing revealed that 33% of OpenURL requests failed. Trainor and Price broke that down further to determine which component in the OpenURL linking chain created the failures. They estimate that a third of those failures were caused by incomplete or inaccurate metadata in the source OpenURL; a third of those failures were caused by knowledge base inaccuracies or translation errors; and a third of those failures were caused because the target mishandled the request coming from the link resolver. If we extrapolate from their sample, it appears that across libraries worldwide about 1 million OpenURL requests fail each day.

The Trainor and Price research provides us with a framework for describing the efforts that are underway to address the errors in OpenURL linking. The problem of uneven OpenURL metadata quality is being addressed through the NISO Improving OpenURLs Through Analytics (IOTA) initiative, described in more detail below. Inaccurate knowledge base content in the link resolver is being addressed by the KBART initiative, described elsewhere in this issue and at <http://www.niso.org/workrooms/kbart>. Improving the way targets handle the incoming request from the link resolver is so far unaddressed formally, but there are discussions underway between members of IOTA and KBART groups to fill that gap.

## **L'Année philologique OpenURL Experiment: Precursor to IOTA**

As Trainor and Price conclusions indicate, the quality of the source OpenURL is critical, but only one of the problems inherent in the OpenURL reference linking model. One of the reasons for the metadata quality problems is a lack of feedback to the source of the OpenURLs. Within vendor organizations that offer OpenURL links in their user interfaces, there is probably some span of management distance between the engineers who add OpenURL functionality to the product and the product manager responsible for the service. The typical OpenURL implementation probably went something like this over the past few years.

1. OpenURL enters the professional discourse
2. Librarians ask if product x is "OpenURL compliant"
3. Product manager submits enhancement request
4. OpenURL enhancement is added to the product roadmap
5. Software engineer does his best, in isolation, to understand the z39.88 standard specification
6. OpenURL feature is rolled out to customers
7. Software engineer moves on to the next project

This was the pattern for L'Année philologique (<http://www.annee-philologique.com/aph/>), an abstracting and indexing database that covers the classics, but following some complaints in 2008, Eric Rebillard, Professor of History and Classics at Cornell

University and editor of L'Année really wanted to understand why the OpenURLs clicked on by users were not always working.

From the perspective of the content provider sending out OpenURLs, actually, it is not possible in practice to determine how well links work, without extensive involvement of a librarian. While the link resolver menu page might display to somebody working at the OpenURL source there is no way to test the vendor's own links from that point outward because authentication (to the "appropriate copy") serves as a gate. This limitation is inherent within the OpenURL reference linking model: links cannot be directly tested for quality by the source.

Professor Rebillard, on behalf of L'Année philologique, approached the Cornell University Library for help with this problem. A colleague in the library, David Ruddy, and I started looking into the problem as a part of a wider investigation that also included canonical citation linking (see <http://cwkb.org> for more about that project). A generous planning grant from the Mellon Foundation made the work possible. A review of literature on metadata quality led us to Hughes' work (2004) developing metadata quality metrics for Dublin Core OAI repositories. Hughes sought to improve the metadata ingested by the Open Language Archives Community (OLAC) open archives repository, so he developed a method for rating incoming metadata records and aggregating that rating up to the data source itself. Building on that work, we analyzed 800,000 OpenURLs and made recommendations about how to improve the metadata in the L'Année philologique OpenURLs. During the experiment we achieved an insight about the dynamic OpenURL reference linking model. The dynamic reference linking

was a response to limitations of bidirectional linking, but the OpenURL solution to reference linking is incomplete. One of the missing components in dynamic reference linking is feedback to OpenURL providers. The planning grant proved the need for a service and the potential for a version of the Hughes metadata evaluation model adapted for OpenURL. See Chandler (2009) and the forthcoming article "Transparent and Scalable OpenURL Quality Metrics" in a spring 2011 issue of D-Lib.

### **Improving OpenURLs Through Analytics (IOTA)**

The proposal to create a NISO working group was approved by the Business Information Topic Committee (<http://www.niso.org/topics/businfo>) on Tuesday, December 8, 2009. The name of the initiative, IOTA, coined by Cynthia Hodgson from NISO, specifically says "Improving OpenURLs," plural, because the focus of our investigation is OpenURL data. We are trying to make the OpenURL reference linking system more precise by improving the inputs. In a nutshell, the IOTA Working Group is developing a suite of tools that any content platform product manager can use to see what is being sent out to customers and compare it to what other vendors are sending out to their customers. Creating tools that will promote a trend towards higher quality, more predictable metadata in OpenURLs will help link resolver vendors improve the quality of experience for patrons.

The NISO IOTA Working Group roster is comprised of a talented group of librarians and vendors, all working together to build a suite of tools so content providers can improve the quality of their OpenURLs. Our roster includes Rafal Kasprowski, Electronic Resources Librarian,

Rice University; Susan Marcin, Licensed Electronic Resources Librarian, Columbia University; Oliver Pesch, Chief Strategist, E-Resource Access and Management Services, EBSCO Information Services; Ellen Rotenberg, Manager, Product Development, Thomson Reuters; Clara Ruttenberg, Electronic Resources Librarian, University of Maryland; Maria Stanton, Director of Content Operations, Serials Solutions; Elizabeth Winter, Electronic Resources Coordinator, Georgia Tech Library; and Jim Wismer, Manager, Software Engineering, Thomson Reuters. Karen Wetzel, Standards Program Manager, NISO, is instrumental in helping us to move work along.

INSERT FIGURE 1: IOTA Report: Occurrence of spage element across OpenURL sources

We have ingested into our repository over 9 million OpenURLs from dozens of different OpenURL providers. Ingesting OpenURL log files is messy. The log files require some amount of preprocessing before we can run them through our metric parser. Each file is pegged to the quarterly time period in which it was received by the link resolver. This temporal dimension makes it possible to monitor changes in quality over time. I wrote the first version of the parser and user interface and installed them on my personal server. After the NISO working group was created the code was migrated to a NISO server ([openurlquality.niso.org](http://openurlquality.niso.org)). Jim Wismer rewrote and improved the parser. I maintain the user interface.

At the time of this writing we offer two descriptive report types, each of which allows an OpenURL provider to compare their OpenURLs against their peers. One report shows all the metrics for a single OpenURL source, and the other shows all the OpenURL sources across one metric. For example, the screenshot in Figure 1 shows how often one of the most critical OpenURL elements, spage (start page), is present across a set of vendors during a time period. We are researching a third report type that we are calling a "completeness index". The completeness index will give each OpenURL source a single rating. We hope such a rating will provide OpenURL providers with a clearer overall picture of how their OpenURLs compare to others'.

In addition to the reports, we are creating documentation, including screencasts showing how to generate custom reports, and case studies to help librarians or vendors understand how to make use of the reports to improve service.

## **Next Steps**

Within the limitations of the OpenURL model, improving the quality of the data flowing into link resolvers is the most effective method to decrease the unacceptable rate of request failures experienced by users every day. The change in the late 1990s from static, bilateral to dynamic reference linking shifted the burden of linking away from the source, where at the time it was overloaded, to the link resolver, where I would argue, it is overloaded today. That is, the link resolver is expected to do too much, and much of what it is expected to do is actually out of its reach to address systematically. What we now know is that OpenURL 1.0 was a first



order approximation of a solution to the appropriate copy problem. The work of IOTA and KBART attempts to uncover and systematically address the second order problems inherent in the OpenURL model by (a) improving the quality of the OpenURL metadata sent to the link resolver (by building into the model a feedback layer) and (b) improving the quality of the holdings data knowledge base used by the link resolver.

There is a third problem in the OpenURL model, alluded to earlier, that needs to be confronted: the continued use of proprietary target link-to syntaxes and behaviors. There has been essentially no change in the ad hoc way that systems handle link resolver requests since the first dynamic reference linking experiments in the 1990s. Back then the solution was clever and resourceful. Now it is an anachronism, a dirty secret. Even today link resolver/knowledge base vendors scramble to track down the syntax of the targets and cross their fingers that the vendor does not change it, just like they did 10 years ago. Each vendor maintains a near duplicate registry of mappings to proprietary syntax links. The vendor syntaxes may change without warning. To compound the problem, the link handling at the target side is idiosyncratic and unpredictable. At Cornell we have observed, for example, that some links will actually fail when more complete metadata, such as an author's last name, is included in a request for full text. This chain in the OpenURL model is overdue for standardization. All parties stand to benefit: patrons (better service), link resolver vendors (better product at less cost), and content providers (more usage). Working group members in IOTA and KBART are currently discussing a joint project to address the gap in the standards landscape.

Links:

Reports: <http://openurlquality.niso.org/>

Blog: <http://openurlquality.blogspot.com/>

Twitter: @nisoiota

### **SIDEBAR: Implications for content providers, link resolver vendors, and librarians**

*Why content providers should care about the quality of your OpenURL:* Your competitors care. Until IOTA, OpenURLs varied in quality, but nobody had a tool to quantify the difference. Now we have the basic descriptive tools to make these distinctions. What was previously a black box is now completely open and transparent.

*Why librarians should care about the quality of OpenURLs:* Your patrons care, and you want them to have the best experience possible. In 2011 we will begin to see OpenURL quality enter into negotiations for new products or renewals. Understanding how a given content provider's OpenURLs compare to other options should be a core competence for librarians making these buying decisions.

*Why link resolver vendors should care about the link resolver experience:* The high failure rate will generate alternative linking solutions that bypass the OpenURL link resolver. Pubget.com, for example, is a direct response to the authentication and linking frustrations of link resolver users (Jones and Connor, 2010).

## References

Beit-Arie, Oren, et al. 2001. Linking to the Appropriate Copy: Report of a DOI-Based Prototype.

D-Lib Magazine. 7(9). <http://www.dlib.org/dlib/september01/caplan/09caplan.html>.

Chandler, Adam, Glen Wiley and Jim LeBlanc. Towards Transparent and Scalable OpenURL Quality Metrics. D-Lib Magazine (forthcoming).

Chandler, Adam. 2009. Results of L'Année philologique online OpenURL Quality Investigation: Mellon Planning Grant Final Report.

<http://metadata.library.cornell.edu/oq/files/200902%20lannee-mellonreport-openurlquality-final.pdf>.

Culling, James. 2007. Link Resolvers and the Serials Supply Chain: Final Project Report for UKSG, p. 33. [http://www.uksg.org/sites/uksg.org/files/uksg\\_link\\_resolvers\\_final\\_report.pdf](http://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf).

HangingTogether Blog. 2009. Herbert's Adventures in Linking." Posted February 5, 2009.

<http://hangingtogether.org/?p=616>.

Hughes, Baden. 2004. Metadata Quality Evaluation: Experience from the Open Language Archives Community. In *Digital Libraries: International Collaboration and Cross-Fertilization*. Ed. Zhaoneng Chen et al. Berlin: Springer-Verlag, 2004, pp. 320-329.

<http://books.google.com/books?id=ixvyvwu->

[1ZIC&lpg=PA320&ots=Wm8h8ikxZH&dq=Hughes%2C%20Baden.%202004.%20Metadata%20Quality%20Evaluation&pg=PA320#v=onepage&q=Hughes%2C%20Baden.%202004.%20Metadata%20Quality%20Evaluation&f=false](http://books.google.com/books?id=ixvyvwu-1ZIC&lpg=PA320&ots=Wm8h8ikxZH&dq=Hughes%2C%20Baden.%202004.%20Metadata%20Quality%20Evaluation&pg=PA320#v=onepage&q=Hughes%2C%20Baden.%202004.%20Metadata%20Quality%20Evaluation&f=false).

Jones, Ryan, and Ian Connor. Telephone conversation, December 7, 2010.

Lynch, Clifford A. 1997. Building the Infrastructure of Resource Sharing: Union Catalogs, Distributed Search, and Cross-Database Linkage. *Library Trends* 45(3), pp. 448-461.

National Information Standards Organization. 2004. ANSI / NISO Z39.88: The OpenURL Framework for Context-Sensitive Services.

[http://www.niso.org/kst/reports/standards?step=2&project\\_key=d5320409c5160be4697dc046613f71b9a773cd9e](http://www.niso.org/kst/reports/standards?step=2&project_key=d5320409c5160be4697dc046613f71b9a773cd9e).

Trainor, Cindi and Jason Price. 2010. Digging into the Data: Exposing the Causes of Resolver Failure. Library Technology Reports 46(7), pp. 15-26.

Van de Sompel, Herbert and Hochstenbach, Patrick. 1999a. Reference Linking in a Hybrid Library Environment. Part 1: Frameworks for Linking. D-Lib Magazine. 5(4).  
[http://www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt1.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html).

Van de Sompel, Herbert and Hochstenbach, Patrick. 1999b. Reference Linking in a Hybrid Library Environment. Part 2: SFX, A Generic Linking Solution. D-Lib Magazine. 5(4).  
[http://www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt2.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html).

Van de Sompel, Herbert and Beit-Arie, Oren. 2001. Open Linking in the Scholarly Information Environment Using the OpenURL Framework. D-Lib Magazine. 7(3).  
<http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html>.

FIGURE 1: IOTA Report: Occurrence of spage element across OpenURL sources

