

Purdue University
Purdue e-Pubs

Charleston Library Conference

NISO's IOTA Working Group: Creating an Index for Measuring the Quality of OpenURL Links

Rafal Kasprowski
Rice University, kasprowski@rice.edu

Susan Marcin
Columbia University, smarcin@columbia.edu

Follow this and additional works at: <https://docs.lib.purdue.edu/charleston>

An indexed, print copy of the Proceedings is also available for purchase at:

<http://www.thepress.purdue.edu/series/charleston>.

You may also be interested in the new series, Charleston Insights in Library, Archival, and Information Sciences. Find out more at: <http://www.thepress.purdue.edu/series/charleston-insights-library-archival-and-information-sciences>.

Rafal Kasprowski and Susan Marcin, "NISO's IOTA Working Group: Creating an Index for Measuring the Quality of OpenURL Links" (2010). *Proceedings of the Charleston Library Conference*.
<http://dx.doi.org/10.5703/1288284314855>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

NISO'S IOTA WORKING GROUP: CREATING AN INDEX FOR MEASURING THE QUALITY OF OPENURL LINKS

Rafal Kasprowski (kasprowski@rice.edu) *Electronic Resources Librarian, Fondren Library
Rice University*

Susan Marcin (smarcin@columbia.edu) *Licensed Electronic Resources Librarian Continuing &
Electronic Resources Management Division, Columbia University Libraries*

ABSTRACT

The IOTA (Improving OpenURLs Through Analytics) working group has been formed by NISO (National Information Standards Organization) in 2010 to develop a community-recognized index for measuring the quality of OpenURL links generated by content providers. OpenURL links do not work as expected all too frequently, and although the OpenURL standard was introduced a decade ago, no systematic method has been implemented to benchmark the resulting links. This paper's objective is to help users better understand what causes OpenURLs to fail and to familiarize them with the OpenURL quality index, a scoring system in development meant to help improve OpenURL linking by measuring the quality of citation source OpenURLs. After an overview of OpenURL and related linking problems, IOTA's work on the OpenURL quality index is discussed in detail. An OpenURL analysis tool is already available for use. It generates reports that compare OpenURLs across databases and vendors, which can help OpenURL providers improve their links. Community members may follow IOTA's progress on the group's web sites and support the initiative by contributing their OpenURL log files.

FULL-TEXT LINKING BEFORE OPENURL

Full-text linking did not begin with the advent of the OpenURL standard ten years ago. Before OpenURL linking was available, proprietary full-text linking was the only option. It continues as an alternative and complement to OpenURL linking today. Contrary to OpenURL linking, proprietary linking does not rely on a standard linking syntax, but on multiple linking syntaxes each dictated by an individual content provider.

Traditionally, certain providers offered proprietary full-text linking as a content enriching feature to their abstracting and indexing (A&I) databases (e.g., CSA's Illumina platform and PubMed.gov). These A&I providers would form agreements with article content providers to make full-text linking available to their users. Users were then able to manually select the full-text providers they had subscriptions with through the administrative modules of compliant A&I databases, thereby establishing links leading from the citations in the A&I database to full text articles they had licensed for those citations. In this full-text setup, the burden of the work was on the A&I providers, who had to create the online administrative modules, and the customers, who had to make the selections appropriate to their holdings. A&I providers offering this type of full-text linking were few and the number of full-text providers they offered links to was relatively limited. Once the setup to a full-text provider was completed, however, the links lead to the corresponding articles fairly reliably.

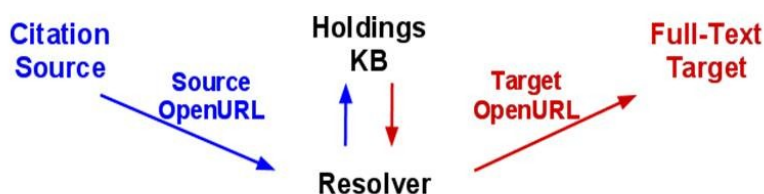
OPENURL FULL-TEXT LINKING

OpenURL introduced a significant shift in the full-text linking model. The goal of its developers was to introduce an open standard that freed the delivery of full texts from proprietary silos. The newly created standard allowed A&I providers and full-text providers to exchange citation metadata between each other. The library's holdings would be provided by its journal manager (the "A-Z list"). The holdings information fed into the journal manager now constituted the knowledge base that would determine the "appropriate copy" to be delivered to users based on the library's subscriptions. As vendors of journal management systems began to use OpenURL linking as a selling point for their products, the burden on A&I providers to develop an administrative module for setting up the full-text linking was lifted. This paved the way for a broad adoption of the OpenURL standard by the library community.

The central element of the OpenURL linking system is the link resolver, a service that ensures the flow of the citation metadata within an URL structure from the A&I provider through the A-Z list to the full-text provider following a defined linking syntax (see Figure 1). The link resolver checks the data from the A&I provider (or "source") against the library's holdings in its journal manager (or "knowledge base") and, based on that information, fetches the item for the user from the appropriate full-text provider (or "target"). As part of the OpenURL setup, active links are located next to the citations in the A&I resources. Clicking these links triggers the link resolver.

Two other features distinguish OpenURL linking from proprietary linking. First, OpenURL links are only generated when a user requests a full text; second, the final link is only completed once the citation metadata reaches the full-text provider. Unlike proprietary links, which are predefined and static, OpenURL links are generated as needed and formed dynamically, following a standard syntax.

OpenURL Syntax and Resolver



Source Citation

A, Bernard, et al. "A versatile nanotechnology to connect individual nano-objects for the fabrication of hybrid single-electron devices." *Nanotechnology* 21, no. 44 (November 5, 2010): 445201. *Academic Search Complete*, EBSCOhost (accessed October 24, 2010).

Target OpenURL (Source OpenURL is structured similarly)

<http://ps4ps6lm2r.search.serialssolutions.com/?issn=0957-4484&volume=21&issue=44&date=20101105&spage=445201&title=Nanotechnology&atitle=A+versatile+nanotechnology+to+connect+individual+nano-objects+for+the+fabrication+of+hybrid+single-electron+devices.&aulast=A++Bernard>

Figure 1. The link resolver passes the citation metadata for the requested full text inside an URL to check for the institutional holdings in a knowledge base (KB) and then, based on the holdings, to communicate with the appropriate provider to retrieve the full text.

PROBLEMS WITH OPENURL LINKING

Although OpenURL has become the more prevalent full-text linking model, it is also the less predictable of the two. Distributed over several link nodes (A&I provider, resolver, journal manager, and full-text provider), OpenURL linking both multiplies the possible locations of link failures and makes them more difficult to pinpoint. OpenURL linking failures are also not uncommon. A recent study reported that "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)." (Culling 2007, 33) Despite its weaknesses, OpenURL linking has not improved significantly since its inception ten years ago, and no systematic method has been developed to benchmark OpenURLs.

Current metadata research, however, has brought forth concepts that may lead to improvements in OpenURL linking. One study, for example, indicates the need for metadata quality metrics, such as completeness, accuracy, conformance to expectations, logical consistency and coherence. (Bruce and Hillmann 2004) Another proposes a method for evaluating metadata quality that would lead to the creation of a baseline for comparison, result sharing between data providers, and the development of a best practice. (Hughes 2004)

OPENURL STUDY AT CORNELL UNIVERSITY

In 2008, some of these new concepts were applied in a Cornell University study led by Adam Chandler and focusing on the OpenURL links associated with the *Année philologique* citation resource. (Chandler, Results of L'Année philologique online OpenURL Quality Investigation: Mellon Planning Grant Final Report 2009) The OpenURLs generated by this database were found to have a high failure rate in resolving to the requested full-text resource. The objective of the Cornell study was to examine the quality of these OpenURLs in a first step towards improving them.

The problem analysis in the Cornell study limited itself to source links and the presence or absence of citation metadata elements. It proposed an OpenURL quality model that compared the OpenURL elements in *Année philologique* to those of other OpenURL providers. It came short of establishing the intended scoring system for measuring the quality of OpenURLs, but the suggested OpenURL quality model was a first step in that direction.

FORMATION OF IOTA

Encouraged by the initial results of the *Année philologique* study, Chandler submitted a proposal to NISO to pursue the research begun at Cornell University and present its results to a wider community. The proposal was accepted by NISO, which led to the formation, in January 2010, of the OpenURL Quality Metrics working group, renamed the Improving OpenURLs Through Analytics (IOTA) working group later that year.

IOTA's approach follows these basic precepts:

1. Research results are obtained through an analytical investigation of how OpenURL links work.
2. The aim is to resolve practical problems; the quality of OpenURL links is addressed, not the underlying standard.
3. Significant improvements in OpenURLs can be achieved by selective changes to the links. (This perspective is reflected in IOTA's motto: "Small changes, big improvements".)

IOTA's objectives are to:

1. develop a community-recognized index for measuring the quality of OpenURL links generated by content providers;
2. produce qualitative reports that will help OpenURL providers quickly compare the quality of their links with those of their peers;
3. use work methods that are fair, transparent, and scalable across all OpenURLs and their providers.

OPENURL COMPLETENESS INDEX

To begin analyzing the citation metadata, IOTA immediately began to collect OpenURLs from libraries and content providers, reaching the 9 million mark by the end of 2010. An OpenURL parsing tool, based on the OpenURL quality model first proposed in the Cornell study, compares the frequencies of citation metadata elements between OpenURL providers, and provides the framework for an index that will ascribe quality scores to those providers' OpenURL sets.

The scoring system is based on the assumption that a correlation exists between the number of core elements in the OpenURLs and the ability of the links to resolve to the desired content. The core elements are any of the citation metadata elements found in the collected OpenURLs. The more core elements an OpenURL contains the more complete it is considered to be; hence the designation of "OpenURL completeness index" for this method of attributing a score to OpenURLs. The OpenURL completeness index draws not only on the number of core elements in the link, but also on their relative weighting. Core elements considered to contribute more to the overall ability of an OpenURL to resolve to a full text receive a higher weighting. For example, an ISSN would be considered a more important element than a journal title, a starting page number more than an article title, a date more than any of the preceding elements, while DOIs or PMIDs would have the highest weighting, as the single presence of one of these identifiers may be sufficient for a successful link.

An initial formula for the scores has been proposed (see Figure 2). A provider's rating would be based on the ratio between the total possible points that can be obtained for the OpenURLs collected for that provider (denominator) and the actual points obtained (numerator). The total possible points would be the product of the number of OpenURLs collected for that provider and the number obtained by adding the weighted points of each different metadata element collected for all the providers. The actual number of points obtained is the cumulative weighting of all the elements found for that one provider.

OpenURLs in sample	112		
	count (# of OpenURLs in sample (112) that contain this element)	weighted points	count * weighted points
artnum	0	1	0
atitle	72	1	72
aulast	88	1	88
btitle	0	1	0
date	107	5	535
doi	87	8	696
eissn	23	3	69
isbn	0	3	0
issn	78	3	234
issue	69	3	207
jtitle	96	1	96
pmid	80	8	640
spage	107	3	321
title	0	1	0
volume	107	3	321
		45	3279
total possible points = OpenURLs in sample * total weighted points for all elements (B5*C24):		5040	
rating = points for provider during period / total possible points for provider during period (D24/C28):		0.65	

Figure 2: Total possible points = OpenURLs in sample * total weighted points for all elements = 112*45 = 5040

Rating = points for provider during period / total possible points for provider during period = 3279/5040 = 0.65

(Chandler, LeBlanc and Wiley, Transparent and Scalable OpenURL Quality Metrics forthcoming)

The scoring system remains a work in progress as a better understanding of the OpenURL completeness concept has revealed weaknesses in the original version of the index. Determining a satisfactory relative weighting for each element, in particular for the identifiers, will require more testing. Since most providers still do not offer content in a variety of formats, a single scoring system may introduce bias in the results. For example, it is obvious from the lack of the "btitle", "isbn", and "title" elements in Figure 2 that the OpenURL provider does not carry any materials in a format that would require their presence, such as books. Currently a project is underway to separate the elements found in journal OpenURLs from those found in OpenURLs for books and book chapters (using the genre field for example). Once this work is completed, the user interface will present a format selection parameter for journals, books, or both. Additional work on preprocessing the OpenURL data received from the community is also expected to help identify alike data, reduce noise, and produce more consistent results.

IOTA is aware that problems with citation metadata are not limited to OpenURLs from citation sources. OpenURLs can, for example, also present problems at the target level, where the citation metadata is exchanged with the full-text provider. In fact, a high degree of "completeness" for an OpenURL is not always indicative of a high level of "success" in linking to full texts. A combination of multiple indexes, which would reflect linking activity along each of the link nodes, could lead to a more comprehensive analysis. While additional OpenURL quality indexes may provide valid results in the future, the OpenURL completeness index could become a valuable tool in itself for measuring the quality of OpenURL links, provided that sufficient consistency and accuracy can be achieved in the OpenURL analysis at the citation source level.

COMMUNITY-DERIVED OPENURL REPORTS

While the OpenURL scoring system is being developed, users can already perform qualitative comparisons of the OpenURLs stored using the OpenURL analysis tool on the IOTA site. Two general types of reports can be generated: source-centric or element-centric. In the source-centric reports (i.e. by source vendor or source database), the contents can be analyzed in terms of the frequency a core element has been included in the OpenURL strings for each vendor or database. There are two ways to analyze the core element frequency, at the element level or at the level of the various formats (or "patterns") in which each element can be displayed. The "isbn" element, for instance, is divided into these formats: "isbn_10", "isbn_13", and "isbn_other". In the element-centric reports, one can view how many times an element or element format appears across either all source vendors or all source databases. In both types of reports, the frequencies obtained are compared against the logsource for the time period selected (year and quarter). The logsource represents the contributors of the OpenURL data to IOTA. Users can select individual contributors or all of them as the logsource for any given report (see Figure 3). OpenURL providers that have contributed their OpenURL data can run reports of their own data, compare it to the OpenURL data from other providers, and evaluate where their OpenURLs may need improvement for better linking results. Since the results can be batched quarterly or annually, contributing providers can submit their OpenURLs to the system periodically and monitor their progress over time.

Home > **Select Report Type**

1. Choose log file provider

2. Choose report type

- source report
- element and pattern frequency

all

AIP

The Claremont Colleges

Cornell University

EBSCO Information Services

Georgia Institute of Technology

Hispanic American Periodicals Index (HAPI)

Kansas State University

WorldCat Link Manager

Serials Solutions

Thomson Reuters

atabases

ext



[Z39.88 \(OpenURL 1.0 standard\)](#) | [OpenURL registry](#)

Figure 3: Selecting the logsource in IOTA's OpenURL analysis tool. (IOTA Working Group n.d.)

PROMOTING IOTA'S WORK

Besides the OpenURL log file reports, the IOTA project site also leads to a wiki and news blog. The wiki contains user documentation on the OpenURL reports, the OpenURL processor database structure, and the preprocessing performed on the data, as well as an IOTA FAQ and other relevant information. The news blog reports on IOTA's activities and significant OpenURL issues. OpenURL providers that wish to become IOTA contributors may contact the group by email (see below). The community at large is also encouraged to take advantage of the tools and information presented on the IOTA site and to communicate with colleagues about IOTA and other OpenURL improvement initiatives.

Group email: openurlquality@list.niso.org

NISO IOTA site: <http://www.niso.org/workrooms/openurlquality>

Project site: <http://openurlquality.niso.org/>

Wiki: http://www.assembla.com/wiki/show/oq/UI_documentation

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

Twitter: #nisiota



WORKS CITED

Bruce, Thomas R, and Diane I Hillmann. "The Continuum of Metadata Quality: Defining, Expressing, Exploiting." In *Metadata in Practice*, by Diane I Hillmann and Elaine L Westbrooks, 238-256. Chicago: American Library Association, 2004.

Chandler, Adam. "Results of L'Année philologique online OpenURL Quality Investigation: Mellon Planning Grant Final Report." 2009. <http://metadata.library.cornell.edu/oq/files/200902%20annee-mellonreport-openurlquality-final.pdf> (accessed January 27, 2011).

Chandler, Adam, Jim LeBlanc, and Glen Wiley. "Transparent and Scalable OpenURL Quality Metrics." forthcoming.

Culling, James. "Link Resolvers and the Serials Supply Chain: Final Project Report for UKSG." 2007. http://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf (accessed January 27, 2011).

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

IOTA Working Group. "Select Report Type." *Improving OpenURLs Through Analytics*. <http://openurlquality.niso.org/select.php> (accessed January 27, 2011).