NISO RP-21-2013

Improving OpenURLs Through Analytics (IOTA): Recommendations for Link Resolver Providers

A Recommended Practice of the
National Information Standards Organization

Approved: April 26, 2013

About NISO Recommended Practices

A NISO Recommended Practice is a recommended "best practice" or "guideline" for methods, materials, or practices in order to give guidance to the user. Such documents usually represent a leading edge, exceptional model, or proven industry practice. All elements of Recommended Practices are discretionary and may be used as stated or modified by the user to meet specific needs.

This recommended practice may be revised or withdrawn at any time. For current information on the status of this publication contact the NISO office or visit the NISO website (www.niso.org).

Published by
National Information Standards Organization (NISO)
3600 Clipper Mill Road, Suite 302
Baltimore, MD 21211
www.niso.org

ISBN: 978-1-937522-18-6

Contents

Foreword

About this Recommended Practice

NISO's OpenURL Quality Committee, which later became known as IOTA (Improving OpenURL Through Analytics), was given a charge that contained the following problem statement:

> The OpenURL standard is widely deployed technology to facilitate linking to resources across the library supply chain. The OpenURL-formatted URL carries the data about an item to the link resolver of the library. The resolver compares the metadata embedded within the OpenURL with what is held in the library's collection and presents the available options in a results page. For a book, there is usually a link to the library's catalog card; for an article, ideally this is a link directly to the full text of the article. At a typical academic library, thousands of OpenURL requests are initiated by patrons each week. The problem is that too often these links do not work as expected, leaving patrons frustrated by a lower than desired quality of service. Periodically, mention is made in the library literature of problems with OpenURL linking, but since the OpenURL standard was introduced a decade ago, no systematic method has been designed and carried out to benchmark it. This work is intended to fill the gap.

As part of its work, the committee analyzed millions of OpenURLs and developed the notion of a Completeness Index as a means of quantifying OpenURL quality. The committee found that there was a pattern to the failures in OpenURLs. OpenURLs that included certain data elements performed better than others missing one or more of those elements. The Completeness Index was developed as a method of predicting the success of OpenURLs from a given provider by examining the data elements that provider includes in the OpenURLs from its site. The index acknowledges that certain data elements are more critical to success than others by giving these elements a higher weight.

Technically speaking, a Completeness Index for a given provider is the average Completeness Score for all OpenURLs being analyzed from that provider. The Completeness Score is the sum of the weights for each of the core elements included in the OpenURL divided by the total potential score. If all core elements were included the score would be 1.

The theory behind the Completeness Score and Completeness Index was validated by separate tests where thousands of OpenURLs were tested for successfully creating a link to full text and this success was correlated against the Completeness Score. These independent tests by both EBSCO and Serials Solutions validated the concept. The tests also confirmed a suspicion that creating a single universal set of element weights is not practical. Differences in linking environments and link resolver technologies affect the importance of certain elements. For example, a link resolver that does no enhancements to the data provided in the OpenURL will have a high failure rate if no ISSN is provided; however, the typical commercial OpenURL link resolver can use the journal title to look-up the ISSN and thus the absence of the ISSN is less critical. And to use another example, a link resolver that is able to use volume, issue, author, and article title to look-up an article in a service like CrossRef® will be more forgiving to a missing Start Page in the OpenURL than one that does not offer such article metadata enhancement.

Coming out of the work of the IOTA committee was the recommendation that link resolver providers (or others interested in an OpenURL linking environment) introduce the notion of the Completeness Index with its constituent Completeness Scores to introduce a quantitative mechanism for evaluating link quality from different providers. Because the element weights are environment-dependent (as illustrated in the ISSN and Start Page examples above), the committee has created this Recommended Practice to serve as a guide for: calculating the element weights, generating a Completeness Index, and analyzing the results and affecting change.

NISO Discovery to Delivery Topic Committee Members

The Discovery to Delivery (D2D) Topic Committee had the following members at the time it approved this Recommended Practice.

Tim Babbitt
Cambridge Information Group

Pascal Calarco, Co-chair
University of Waterloo Library

Peter Murray
Lyrasis

Jeff Penka
OCLC Online Computer Library Center

Tim Shearer
University of North Carolina Chapel Hill Libraries

Lucy Harrison, Co-chair
Florida Virtual Campus

Chris Shillum
Reed Elsevier

NISO IOTA Working Group Members

The following individuals served on the NISO Improving OpenURLs Through Analytics (IOTA) Working Group, which developed and approved this Recommended Practice:

Adam Chandler, Chair
Cornell University Library

Rafal Kasprowski
Rice University Fondren Library

Susan Marcin
Columbia University Libraries

Oliver Pesch
EBSCO Information Services

Clara Ruttenberg
University of Maryland Libraries

Elizabeth Winter
Georgia Tech Library

Jim Wismer
Thomson Reuters

Aron Wolf
Serials Solutions

Acknowledgements

The IOTA Working Group wishes to thank Phil Davis for the valuable assistance he provided, guiding us in our stepwise regression method.

Trademarks, Services Marks

Section 1: Introduction

1.1 Purpose and Scope

These recommended practices are intended to assist information professionals involved in managing OpenURL linking environments in improving OpenURL linking by providing them with the tools necessary to measure OpenURL quality.

The focus of the IOTA (Improving OpenURL Through Analytics) Working Group is the quality of the metadata that is passed to the link resolver from the OpenURL source. The quality of the data in the link resolver knowledge base itself is outside the scope of IOTA; this is being addressed through the NISO KBART initiative. Also outside of the scope is the quality of the linkages between link resolvers and full text content providers. (To our knowledge, this part of the OpenURL quality problem is not being addressed by any organized initiative.)

This Recommended Practice is focusing on a specific genre of OpenURLs—those intended to provide access to journal articles. The information provided in this Recommended Practice could easily be adapted for other genres such as book and book chapters.

1.2 Terms and Definitions

The following terms, as used in this recommended practice, have the meanings indicated. Terms in boldface in a definition indicate the term is also defined in this section.

Term	Definition
A&I database Abstract and Index database	A content discovery product that provides descriptive information about content items but not the full text of the content. A&I databases are frequently a **source** in **OpenURL** linking.
completeness	The number of metadata **elements** provided in the **OpenURL** out of a desired or core number.
Completeness Index	A number that is attributed to a content provider (OpenURL r**eferrer** or **source**) to measure the **completeness** of the provider's **OpenURLs** in aggregate. It is essentially an average of the **Completeness Scores** of **OpenURLs** coming from that content provider.
Completeness Score	The measure of the **completeness** of a single **OpenURL**.

Term	Definition
DOI® Digital Object Identifier	A unique and persistent identifier for a digital content object, such as an online article, journal, chapter, book, or image, using a syntax defined in ANSI/NISO Z39.84 and ISO 26324. The DOI provides current information about the object, including where to find it on the internet. Typically registered with an authorized registration agency, e.g., CrossRef (www.crossref.org), whereby it is possible to use the DOI to look up descriptive metadata about a content item using a provided free service. Sometimes called a DOI name.
discovery platform	A website or product through which users can discover content items.
element	Descriptive metadata such as ISSN, Volume, Issue, Start Page, etc. used in an **OpenURL**.
Element Weight	A value assigned to an **element** and used in calculation of a **Completeness Score**. The value of the Element Weight represents the relative importance of the **element** to the success of the **OpenURL** in providing the user access to full text; the higher the value the more important the **element**.
Enhanced OpenURL	The set of **OpenURL elements** available after the **link resolver enhancers** have been run on the incoming **OpenURL**. Typically an Enhanced OpenURL will have more data **elements** than the original **OpenURL**.
enhancer	An automated process that enhances an **OpenURL** by using internal and external sources to supplement the **OpenURL** data **elements**. An example would be an enhancer that looks up article-level metadata from CrossRef using the **DOI** presented on an **OpenURL**.
fail	Describes the state of an **OpenURL** that does not generate any item-level links to full text items. *See also* **success**. NOTE: Within the context of this recommended practice the following are not considered links to full text items: links to a journal homepage, the table of contents for an issue, or a search page at the vendor site.
full text target link	A link to the complete text—including all references, figures, and tables—of an article on the **target** site. If the referenced item is a journal article, the full text article link would direct the user directly to that article at the content provider's site without requiring further navigation or searching.

Term	Definition
knowledge base	A database used with **OpenURLs** that contains information about what **targets** are available to users of the **link resolver**. Data within the knowledge base includes but may not be limited to link syntaxes and holdings with coverage details. The **link resolver** environment for a given institution can be customized to reflect that institution's collection and to only provide links to **targets** to which the library subscribes.
link resolver	Technology that controls the linking between **sources** and **targets**. The link resolver accepts and deconstructs the **OpenURL** describing a content item from a **source** and uses its **knowledge base** and associated programs to determine full text and other t**argets** appropriate for the user and create predictable links to these. The role of the link resolver is context sensitive linking to the appropriate copy of a content item.
OpenURL	As defined in ANSI/NISO Z39.88-2004, *The OpenURL Framework for Context-Sensitive Services*, a URL designed to transport metadata and thus enable linking from information resources such as **abstracting and indexing databases** (**sources**) to library services (t**argets**), such as academic journals, whether online or in printed or other formats. The linking is mediated by **link resolvers**, or link-servers, which parse the elements of an **OpenURL** and provide links to appropriate t**argets** available through a library by the use of a knowledge base. *Source:* Wikipedia **[boldface notation added]**
PubMed ID	A unique identifier assigned to a record in the PubMed databases (www.pubmed.org). A PubMed record typically describes an article or other content item. Using a PubMed ID it is possible to look-up descriptive metadata of the content item through a free online service provided by the National Library of Medicine.
referrer	The identity of a website or **discovery platform** that is the **source** of the **OpenURL**. In an **OpenURL** this is represented as the SID (Source ID) or the Referrer ID. In some contexts, **referrer** is used interchangeably with **source**.
required data elements	The data **elements** that must be available to create a successful link to an **OpenURL** t**arget**. For example, if a publisher's link to a full text article is constructed from ISSN, Volume, Issue and Start Page values, then these fields are considered required data elements since the link will fail to access the full text if any one of them is missing.

Term	Definition
source	The website where the user discovered an item of interest and from which the **OpenURL** was initiated. The **source** is identified by the Referrer ID. In some contexts, **source** is used interchangeably with **referrer**.
success	Describes the state of an **OpenURL** that is able to generate an item-level link to one or more full text items. The link generated is intended to populate the **link resolver** menu; however, the designation of "success" does not guarantee a user will get to the full text item if the link is followed all the way to the full text content provider's site. *See also* **fail**.
Success Score	A value attributed to an **OpenURL** based on its **success** in generating a link to one or more full text items. If the **OpenURL** generates a link to a full text item, it is given a Success Score of 1; if not, the Success Score is 0.
target	The website where the full text resides and where the user will be linked to view the item via the **link resolver**. Example targets could include content in publisher platforms, institutional catalogs or repositories, and content gateways.

Section 2: Determining Optimal Element Weights

This section of the Recommended Practice describes the steps to follow to create Completeness Index Element Weights that have been optimized for a particular linking environment. It starts with the theory behind the recommended approach and continues with guidance on preparing the link resolver environment for testing, selecting OpenURLs to include in the sample, conducting the tests, and, finally, using the results of the test to determine the Element Weights.

2.1 The Stepwise Regression Approach

The Completeness Index can be a good indicator of the probability of an OpenURL's success or its ability to generate a link to the full text of the referenced item. The elements that are used within the OpenURL links are weighted according to their importance for the success of the link. One way to determine the importance of a given element is to observe the effect on the success of OpenURL links if that element is omitted. The more OpenURLs that fail without the element included, the more important that element is to the success of the OpenURLs. The element's importance will be reflected in the value of the weight assigned.

Element weights are determined using stepwise regression and a form of backward elimination in which core elements are removed from OpenURLs to see the effect on their success. The core data elements for journal article linking are defined in Table 1.

Table 1: Core elements for stepwise regression for journal article linking

Element	Description
atitle	Article title
aulast	Author's last name
date	Date of publication
issn	ISSN (either online or print ISSN)
issue	Issue number
spage	Start page
title	Journal title (Note that the element "jtitle" will be treated as the equivalent to "title".)
volume	Volume number

Note that the DOI and PubMed ID are intentionally excluded as core data elements for the stepwise test. Most link resolvers contain enhancers that will use the DOI or PubMed ID to look up the complete set of metadata for the item identified; therefore, an OpenURL with a DOI or a PubMed ID is equivalent to an OpenURL having all core metadata elements. Including the DOI or PubMed as a core element would thus unduly influence the results of the stepwise regression.

The stepwise regression tests requires a set of OpenURLs that have all core OpenURL elements and have been tested to resolve successfully by locating at least one full text target for the item referenced by the OpenURL (see section 2.2 for configuring a test environment). The sample size for such a test

should be greater than 1,000 OpenURLs to ensure there is sufficient distribution of results across the potentially hundreds of possible targets.

Determining the weight of an element starts with finding the percentage of OpenURLs that fail when that element is missing. The test set of OpenURLs will be modified for each element listed in Table 1 by removing that element. Modified OpenURLs will be executed through the link resolver and the number of links that fail will be tallied for each. Table 2 is an example of results derived using a commercial link resolver and a test of 1,500 OpenURLs.

Table 2: Example of regression test results

Element Removed from the OpenURL	Description	Failure Percentage
atitle	Article title	.74%
aulast	Author's last name	.07%
date	Date of publication	.4%
issn	ISSN (either online or print ISSN)	22.02%
issue	Issue number	20.27%
spage	Start page	33.27%
title	Journal title (either title or jtitle accepted)	.61%
volume	Volume number	74.14%

Since the values for failure rates range widely from less than 1 tenth of a percent to 74 percent the element weights are calculated as log(10) of the failure rate per 10,000 OpenURLs[1] to ensure all values are positive with a reasonable variance in weights. Table 3 shows the weights calculated using the test results shown in Table 2.

Table 3: Sample elements weights derived from failure rates

Core Element	Failure Percentage	Element Weight
atitle	0.74%	1.87
auLast	0.07%	0.85
date	0.40%	1.60
issn	22.02%	3.34
issue	20.27%	3.31
spage	33.27%	3.52
title	0.61%	1.79
volume	74.14%	3.87

1 The approach of log(10) of the failure rate per 10,000 OpenURLs was selected after testing several other algorithms including the actual percentage, log(10) of the failure percentage, and square root of the failure percentage. Log(10) of the failure rate per 10,000 produced the highest correlation coefficient when comparing success of the OpenURL (whether or not it produced a full text target link) to the Completeness Scores based on the test set of element weights.

2.2 Preparing the Link Resolver Environment

The process of determining optimal element weights requires the link resolver environment to test a large number of OpenURLs and record whether or not each OpenURL found at least one appropriate full text target.

2.2.1 Automated Testing Using a Batch Process

The number of OpenURLs to be processed to achieve meaningful results will be quite large. Conducting tests on a thousand or more OpenURLs requires a batch process in which a file containing thousands of OpenURLs can be executed and the results automatically recorded. If the link resolver being tested does not have a batch mode, a simple script or application can be written to handle the task. (Refer to Appendix A for an example of how to carry out this process.

The goal of the batch process is simple:

- Read each OpenURLs from an input file or stream.
- Execute the OpenURL against the link resolver.
- Check the results from the link resolver to determine if a full text target was returned (see 2.2.2).
- Record the result. (If a full text target was found, set "success" to be true.)
- Repeat until all OpenURLs are read and tested.
- Save the results.

2.2.2 Checking for Full Text Targets

When the link resolver evaluates an OpenURL, the test process must be able to determine if the OpenURL matched to at least one full text target of the desired type. For example, when creating Element Weights for journal article OpenURLs, the process would look for target links that would lead to a full text article (at the article level).

Most link resolvers use a rules-based process for selecting a target link. A holdings check against the knowledge base determines which target may have the full text for the referenced item; a coverage check determines if the item is available at the target site; an element check is used to determine if there are enough elements on the OpenURL to create the needed link. For the purpose of this test, the following criteria are used to determine if a full text target will be returned:

- The item referenced by the OpenURL must be found in the knowledge base.
- Customer-level holdings/coverage entitlements should **not** be applied during this test.
- The Enhanced OpenURL must include all the required data elements necessary for the full text target to successfully link to the referenced item
- The full text target link must link directly to the referenced item or a landing page for the item. If there is insufficient data to create a link directly to an article, many link resolvers will offer a link to the journal homepage instead; the idea is that the end user can then search for the article on the content site. However, since our test is about measuring the success of the OpenURL based on its ability to produce a direct link to a full text article, links that go to a journal homepage, a generic search screen, or even an issue-level table of contents should not be considered full text links for the purposes of determining Element Weights; nor should links to ILL or Document Delivery forms be considered successful links.

2.3 Selecting OpenURLs for the Stepwise Regression

The stepwise regression analysis requires a set of OpenURLs that include all the core data elements (called perfect OpenURLs) and which resolve to find at least one full text target. Selecting OpenURLs is a two-step process; the first step is to identify a set of OpenURLs with all the core data elements, and the second step is to identify the subset of these that also resolve to find a full text target. In order to have the number of OpenURLs in the second set number in excess of one thousand, the first set may need to be several times larger.

If there is a source of OpenURLs, such as from the logs of the link resolver being tested, that can be used. It is recommended, however, that a file similar to the one in Figure 1 be created and the OpenURLs be parsed so that each element is in a separate column. This will make it easier to conduct the stepwise test.

ID	URL	DateTime	Referrer	Major Referrer	Genre	Atitle	Aulast	Date	Issn	Issue	jTitle	Spage	Volume
5	?genre=article&isbn=&issn=10502556&title=Journal+of+Divorce+%26+Remarriage&volume=49&issue=1&date=20080101&atitle=Conflict+in+divorcing+and+continuously+married+families%3a+A+study+of+marital%2c+parent-	2012-05-29 11:17	EBSCO:PsycINFO	EBSCO	Article	conflict+in+divorcing+and+continuously+married+families%3a+a+study+of+marital%2c+parent-child+and+sibling+relationships.	noller%2c+patricia	20080101	1050-2556	1	journal+of+divorce+%26+remarriage	1	49
6	?sid=HWW:OMNIFT&genre=article&pid=<an>200210502570009</an>&aulast=Shabani&aufirst=Daniel+B.&issn=0021-8855&title=Journal+of+Applied+Behavior+Analysis&stitle=J+Appl+Behav+Anal&atitle=Increasing+social+initiations+in+children+with+autism:+effects+of+a+tactile+prompt&volume=35&issue=1&spage=79&epage=83&date=2002&ssn=spring	2012-05-29 11:17	HWW:OMNIFT	HWW	Article	increasing+social+initiations+in+children+with+autism:+effects+of+a+tactile+prompt	shabani	2002	0021-8855	1	journal+of+applied+behavior+analysis	79	35
10	?genre=article&isbn=&issn=08943796&title=Journal+of+Organizational+Behavior&volume=20&issue=5&date=19990901&atitle=The+costs%2c+benefits%2c+and+limitations+of+organizational+level+stress+interventions.&aulast=Briner%2c+Rob+B.&spage=647&pages=647-664&sid=EBSCO:PsycINFO&pid=%3	2012-05-29 11:17	EBSCO:PsycINFO	EBSCO	Article	the+costs%2c+benefits%2c+and+limitations+of+organizational+level+stress+interventions.	briner%2c+rob+b.	19990901	0894-3796	5	journal+of+organizational+behavior	647	20

Figure 1: Example of file with OpenURLs selected for testing

Once the initial set of OpenURLs is established, process each through the link resolver and indicate if they were successful or not using the criteria described in section 2.2.2.

For the stepwise test, 1,000 or more of the OpenURLs should result in “success” equals true.

2.4 Conducting the Tests

The stepwise regression requires each OpenURL selected for the test (see section 2.3) to be processed against the link resolver many times—once with each of the core elements removed. One simple way of doing this is to create element-specific variations for each OpenURL in the test. Figure 2 shows a sample file where test OpenURLs have been repeated and adjusted for each element to be tested. In this file, three new columns have been added to the original file: the Element being added; the Test OpenURL, which is derived from the individual elements omitting the element that is the subject of the test; and the Success column, which indicates whether or not the OpenURL resolves to full text. As a point of reference, the original OpenURL with all elements is also included (“all” in the Element being tested column). It is advisable to always retest with all elements whenever conducting the stepwise test to ensure nothing has changed in the knowledge base or linking environment while the test was underway.

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Referrer	Major Referrer	Genre	Atitle	Aulast	Date	Issn	Issue	jTitle	Spage	Volume	Element Being Evaluated	Test OpenURL	Success
2	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**all**	**?genre=Article&atitle=commentary+on+mood+and+memory&aulast=bower&date=19870101&issn=0005-7967&issue=6&jtitle=behavior+research+and+therapy&spage=44&volume=25**	TRUE
3	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**atitle**	**?genre=Article&aulast=bower&date=19870101&issn=0005-7967&issue=6&jtitle=behavior+research+and+therapy&spage=44&volume=25**	
4	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**aulast**	**?genre=Article&atitle=commentary+on+mood+and+memory&date=19870101&issn=0005-7967&issue=6&jtitle=behavior+research+and+therapy&spage=44&volume=25**	
5	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**Date**	**?genre=Article&atitle=commentary+on+mood+and+memory&aulast=bower&issn=0005-7967&issue=6&jtitle=behavior+research+and+therapy&spage=44**	
6	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**ISSN**	**?genre=Article&atitle=commentary+on+mood+and+memory&aulast=bower&date=19870101&issue=6&jtitle=behavior+research+and+therapy&spage=44&volume=25**	
7	EBSCO:PsycINFO	EBSCO	Article	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	**issue**	**?genre=Article&atitle=commentary+on+mood+and+memory&aulast=bower&date=19870101&issn=0005-7967&jtitle=behavior+research+and+therapy&spage=44&volume=25**	

Figure 2: Example of OpenURLs that have been modified to test individual elements

If the test was started with 1,000 OpenURLs and is testing for eight core elements, as is the case with a journal article, the resulting batch of OpenURLs to process will be 9,000—the original 1,000 plus eight variations of each OpenURL representing each element.

It is important to note that none of the test OpenURLs includes identifiers like the PubMed ID or DOI. Because of the way most link resolvers work, inclusion of these identifiers would allow the link resolver's Enhancers to replace the omitted data element and thus render the test results meaningless.

Once the list of test OpenURLs has been created, process each through the link resolver and indicate if the OpenURL was successful by setting the Success column to True if a valid full text target was found, or False if it is not found.

2.5 Calculating Optimal Element Weights

With the test complete, the next step is to perform an analysis by element. The pivot table function in Excel is one way to perform such an analysis by element and showing the count of OpenURL by Success value. Figure 3 offers an example of this method.

	A	B	C	D
1	Stepwise Test Success Analysis by Element			
2				
3		Success of OpenURL		
4	Element being tested	FALSE	TRUE	Total
5	all	0	1000	1000
6	atitle	7	993	1000
7	aulast	1	999	1000
8	Date	4	996	1000
9	ISSN	220	780	1000
10	issue	202	798	1000
11	jtitle	6	994	1000
12	spage	332	668	1000
13	volume	741	259	1000

Figure 3: OpenURL stepwise sample test results analysis

The next step is to calculate the failure rate as a percentage. The calculation is simple; for each element the failure rate is the value in the "False" column divided by the value in the "Total" column. Figure 4 shows the table from Figure 3 with the failure rate calculated.

	A	B	C	D	E
1	Stepwise Test Success Analysis by Element				
2					
3		Success of OpenURL			
4	Element being tested	FALSE	TRUE	Total	Failure Rate
5	all	0	1000	1000	0.0%
6	atitle	7	993	1000	0.7%
7	aulast	1	999	1000	0.1%
8	Date	4	996	1000	0.4%
9	ISSN	220	780	1000	22.0%
10	issue	202	798	1000	20.2%
11	jtitle	6	994	1000	0.6%
12	spage	332	668	1000	33.2%
13	volume	741	259	1000	74.1%

Figure 4: OpenURL stepwise sample test results with failure rates

The final step is the calculation of the Element Weights. This Recommended Practice proposes using log(10) of the failure rate per 10,000 OpenURLs. In our example this would be the formula `=LOG10(<failureRate>*10000)` or to use an Excel formula, the Element Weight for `atitle` would be: `=LOG10(E6*10000)`. Figure 5 shows the sample table with the element weights calculated.

	A	B	C	D	E	F
1	Stepwise Test Success Analysis by Element					
2						
3		Success of OpenURL				
4	Element being tested	FALSE	TRUE	Total	Failure Rate	Element weight
5	all	0	1000	1000	0.0%	
6	atitle	7	993	1000	0.7%	1.85
7	aulast	1	999	1000	0.1%	1.00
8	Date	4	996	1000	0.4%	1.60
9	ISSN	220	780	1000	22.0%	3.34
10	issue	202	798	1000	20.2%	3.31
11	jtitle	6	994	1000	0.6%	1.78
12	spage	332	668	1000	33.2%	3.52
13	volume	741	259	1000	74.1%	3.87

Figure 5: OpenURL stepwise sample test results with failure rates and element weights

Section 3: Generating the Completeness Index

3.1 Reviewing the Theory Behind the Completeness Index

The Foreword of this recommended practice describes how the Completeness Index can be used to measure the probability that OpenURLs from a particular source will be successful. It does this by calculating the average Completeness Score for OpenURLs from that provider. The Completeness Score, which represents the probability that the OpenURL will successfully find a link to the item referenced by the OpenURL, is calculated by taking the sum of the element weights for each of the core elements appearing in the OpenURL and dividing that by the maximum possible score. An OpenURL with all core elements will have a Completeness Score of 1.

Section 2: of this Recommended Practice describes how to determine the optimal Element Weights for a given linking environment.

3.2 Selecting OpenURLs to Include in the Calculation

3.2.1 Source

Ideally the source of the OpenURLs used to calculate the Completeness Index will be actual OpenURLs that have been received by the link resolver being tested.

3.2.2 Genre

Only include OpenURLs of the genre for which the Completeness Index is being calculated. For example, if the analysis of OpenURLs is being performed for journal articles, only include OpenURLs that are explicitly designated as, or default to, the genre of Article. The logic used to classify OpenURLs as article, book, or other is available on the IOTA project management and documentation website (www.niso.org/workrooms/openurlquality).

3.2.3 Quantity

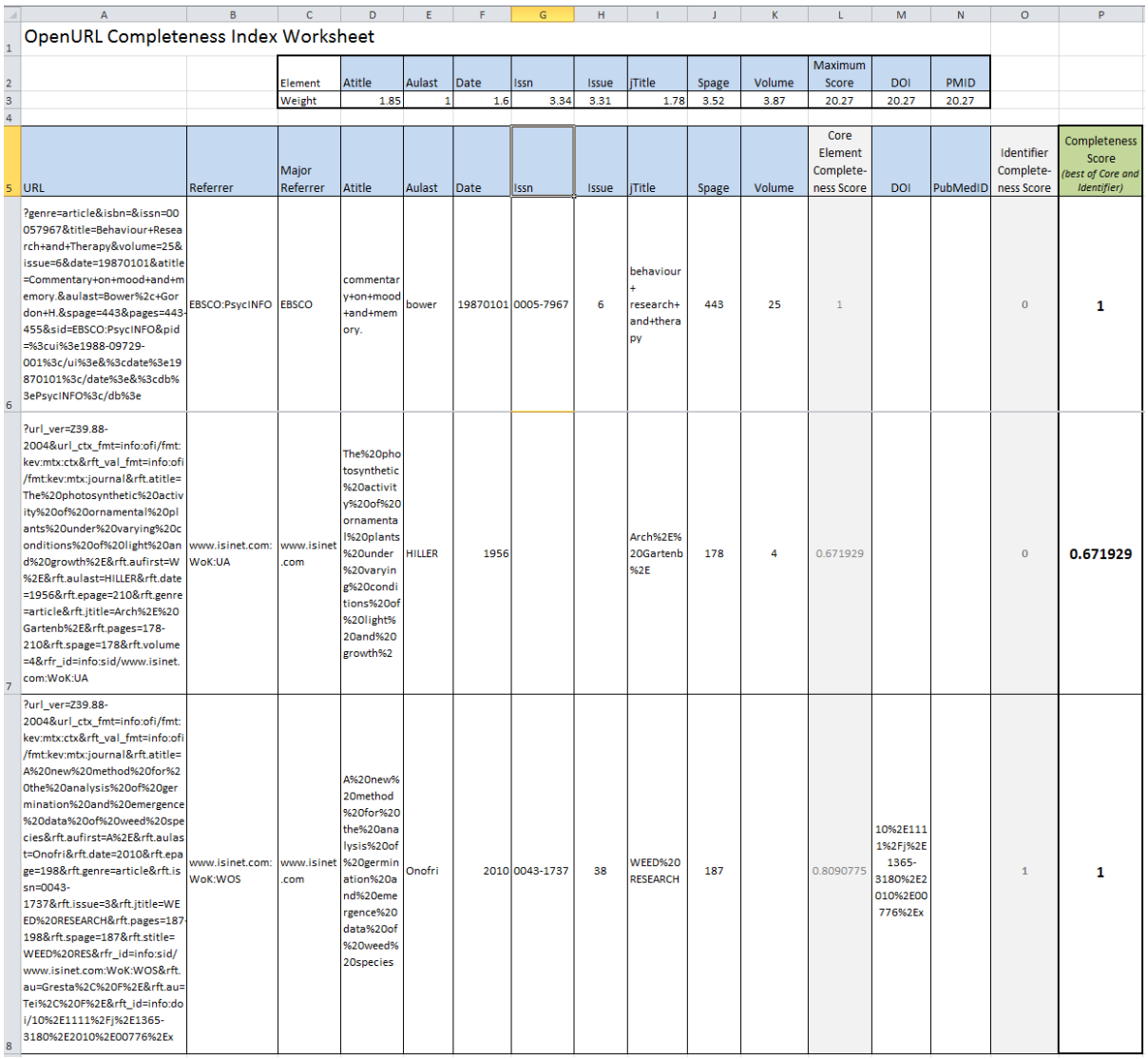
If OpenURLs are selected at random, be sure that the quantity of OpenURLs selected for the test is sufficient so that each of the providers to be included in the test has a sufficient number of OpenURLs to ensure a representative sample.

3.2.4 Time Period

If the goal is to determine the Completeness Index for the current linking environment, it is best to select OpenURLs logged by the link resolver in the past one to three months. If the goal of the exercise is to see how the Completeness Index has changed over time, select multiple sets of OpenURLs from the time periods to be evaluated so that each set will provide a snapshot of completeness from the time period selected, which can then be compared with the others. It is worth mentioning here that the IOTA reporting website (www.openurlquality.org/) is an extremely deep and valuable resource (over 23 million OpenURLs analyzed) for learning about what data is being sent by OpenURL providers over time.

3.3 Preparing for the test

With the OpenURLs selected (see section 3.2), the next step is to prepare data for the calculation. Figure 6 presents an example spreadsheet used to calculate the Completeness Index.



OpenURL Completeness Index Worksheet

	Element	Atitle	Aulast	Date	Issn	Issue	jTitle	Spage	Volume	Maximum Score	DOI	PMID
3	Weight	1.85	1	1.6	3.34	3.31	1.78	3.52	3.87	20.27	20.27	20.27

5	URL	Referrer	Major Referrer	Atitle	Aulast	Date	Issn	Issue	jTitle	Spage	Volume	Core Element Completeness Score	DOI	PubMedID	Identifier Completeness Score	Completeness Score *(best of Core and Identifier)*
6	?genre=article&isbn=&issn=00057967&title=Behaviour+Research+and+Therapy&volume=25&issue=6&date=19870101&atitle=Commentary+on+mood+and+memory.&aulast=Bower%2c+Gordon+H.&spage=443&pages=443-455&sid=EBSCO:PsycINFO&pid=%3cui%3e1988-09729-001%3c/ui%3e&%3cdate%3e19870101%3c/date%3e&%3cdb%3ePsycINFO%3c/db%3e	EBSCO:PsycINFO	EBSCO	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	1			0	**1**
7	?url_ver=Z39.88-2004&url_ctx_fmt=info:ofi/fmt:kev:mtx:ctx&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.atitle=The%20photosynthetic%20activity%20of%20ornamental%20plants%20under%20varying%20conditions%20of%20light%20and%20growth%2E&rft.aufirst=W%2E&rft.aulast=HILLER&rft.date=1956&rft.epage=210&rft.genre=article&rft.jtitle=Arch%2E%20Gartenb%2E&rft.pages=178-210&rft.spage=178&rft.volume=4&rfr_id=info:sid/www.isinet.com:WoK:UA	www.isinet.com:WoK:UA	www.isinet.com	The%20photosynthetic%20activity%20of%20ornamental%20plants%20under%20varying%20conditions%20of%20light%20and%20growth%2	HILLER	1956			Arch%2E%20Gartenb%2E	178	4	0.671929			0	**0.671929**
8	?url_ver=Z39.88-2004&url_ctx_fmt=info:ofi/fmt:kev:mtx:ctx&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&rft.atitle=A%20new%20method%20for%20the%20analysis%20of%20germination%20and%20emergence%20data%20of%20weed%20species&rft.aufirst=A%2E&rft.aulast=Onofri&rft.date=2010&rft.epage=198&rft.genre=article&rft.issn=0043-1737&rft.issue=3&rft.jtitle=WEED%20RESEARCH&rft.pages=187-198&rft.spage=187&rft.stitle=WEED%20RES&rfr_id=info:sid/www.isinet.com:WoK:WOS&rft.au=Gresta%2C%20F%2E&rft.au=Tei%2C%20F%2E&rft_id=info:doi/10%2E1111%2Fj%2E1365-3180%2E2010%2E00776%2Ex	www.isinet.com:WoK:WOS	www.isinet.com	A%20new%20method%20for%20the%20analysis%20of%20germination%20and%20emergence%20data%20of%20weed%20species	Onofri	2010	0043-1737	38	WEED%20RESEARCH	187		0.8090775	10%2E1111%2Fj%2E1365-3180%2E2010%2E00776%2Ex		1	**1**

Figure 6: Example of a Completeness Index Calculation Spreadsheet

The worksheet represented by Figure 6 can automatically calculate the Completeness Scores for individual OpenURLs. The Element Weights are included in a small table in rows 2 and 3. Notice how the element names line up with the element names in row 5, as this is important.

Row 5 contains the header row for the OpenURLs to be analyzed. Table 4 summarizes each column in the spreadsheet.

Table 4: Figure 6 header column descriptions

Column	Label	Description of contents
A	url	The query string of the OpenURLs.
B	referrer	The actual referrer of the content provider extracted from the SID or referrer of the OpenURL.
C	major referrer	The first section of the referrer. For example, if the referrer is EBSCO:PsycInfo, the Major Referrer is “EBSCO”.

Column	Label	Description of contents
D	atitle	The value for the *atitle* element from the OpenURL in column A. If the *atitle* element is missing or empty, this cell will be empty.
E	aulast	The value of the *aulast* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
F	date	The value of the *date* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
G	issn	The value of either the *issn* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
H	issue	The value of the *issue* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
I	jtitle	The value of the *jtitle* or *title* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty. Note that *jtitle* and *title* are used synonymously.
J	apage	The value of the *spage* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
K	volume	The value of the *volume* element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
L	core element complete-ness score	The Completeness Score calculated for the OpenURL in column A. The calculation works by adding the weights in row 3 for any element with a value from the OpenURL being checked. This sum is then divided by the *Maximum Score*. The Excel formula for this is as follows. (This example is for row 6.) `=(IF(D6<>"",D$3,0)+IF(E6<>"",E$3,0)+IF(F6<>"",F$3,0)+IF(G6<>"",G$3,0)+IF(H6<>"",H$3,0)+IF(I6<>"",I$3,0)+IF(J6<>"",J$3,0)+IF(K6<>"",K$3,0))/$L$3`
M	doi	The value of the DOI element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
N	pmid	The value of the PMID (PubMed ID) element from the OpenURL in column A. If the element is missing or empty, this cell will be empty.
O	identifier complete-ness score	The Completeness Score calculated for the identifier. This score will be 1 if either the *DOI* or the *PMID* exist in the OpenURL, or it will be 0 if both are absent or empty. The Excel formula for this follows. (This example is for row 6.) `=IF(M6 &N6 <> "",1,0)`
P	complete-ness score	The final Completeness Score for the OpenURL will be either the Core Element Completeness Score or the Identifier Completeness Score, whichever value is greater. The Excel formula to calculate this is as follows. (This example is for row 6.) `=IF(O6>L6,O6,L6)`

3.4 Calculate Completeness Scores

If a spreadsheet like the one shown in section 3.2 has been successfully created, calculating the completeness score involves the following steps:

- Copying the OpenURLs to be analyzed into column A of that spreadsheet, starting with row 6.
- Copying the element values from the OpenURLs into columns B through K and M and N.
- Adding the formula to column L. Enter the relevant formula from Table 4 into L6 then copy and paste to all rows.
- Adding the formulas for columns O and P. Enter the relevant formulas from Table 4 into O6 and P6 respectively and copy to all rows.

Column P will then contain the Completeness Scores for the OpenURLs being analyzed.

3.5 Generate Completeness Index

Taking the spreadsheet created from section 3.4, use the Excel Pivot Table feature to determine the average Completeness Score for each Major Referrer. Figure 7 shows an example analysis created using an Excel Pivot Table.

	A	B	C
1	OpenURL Completeness Index		
2			
3	Major Referrer	Count of OpenURLs Analyzed	Completeness Index
4	achs	3	1.000
5	ams	40	0.930
6	aph	3	0.946
7	brepols	24	0.864
8	cas	140	0.869
9	csa	85	0.980
10	EBSCO	3,486	0.982
11	hapi	1	0.809
12	hww	44	0.870
13	isinet.com	4,982	0.990
14	itergateway.org	2	1.000
15	libx	52	0.985
16	mendeley.com%2fmendeley	6	1.000
17	muse	5	0.951
18	oclc.org	29	0.947
19	refworks	564	0.876
20	serialssolutions.com	89	0.960
21	umich.edu	1	1.000
22	unknown	1	0.623
23	webfeat	8	0.959
24	worldcat.org	5,438	0.964
25	**Grand Total**	**15,003**	**0.972**

Figure 7: Example Completeness Index

In Figure 7, the Completeness Index is the average of the Completeness Scores for all the OpenURLs analyzed from this provider (as noted in the Major Referrer column). It is also good to include a count of the OpenURLs analyzed in the final Completeness Index. As seen in this example, several of the providers had only had one or two OpenURLs analyzed; therefore, there is not a large enough sample to draw any conclusions.

If there are two few OpenURLs from some key content providers, the sample size will need to be increased by adding more OpenURLs from these content providers.

Section 4: Analyzing Results and Affecting Change

4.1 Interpreting Completeness Index

The Completeness Indexes can serve as a tool to help determine which content providers are more likely to cause linking problems due to missing data elements in their OpenURLs.

4.1.1 Meaning of the Completeness Index Value

In the studies that were done in developing the approach to the Completeness Index, a large number of links were tested for success. When an OpenURL generated a full text target link it was awarded a Success Score of 1 and if no full text link was found the Success Score was 0. All these zeros and ones were added up and then divided by the total number of OpenURLs from the content provider. The average Success Score for a content provider had a very high correlation to the Completeness Index for the same content provider. Based on this, one could make a somewhat loose assumption that the Completeness Index provides a rough ratio of the OpenURLs that are going to generate full text links. For example, a Completeness Index of .75 means that roughly 75% of the OpenURLs are capable of generating full text links.

Interpreting the Completeness Index in this manner is a simple way of understanding the quality of OpenURLs from a given source.

4.1.2 Using the Completeness Index

The Completeness Index can be used proactively by reviewing all of an organization's content providers and focusing on those with lower Completeness Index values in an attempt to improve the quality of OpenURL coming from those sites.

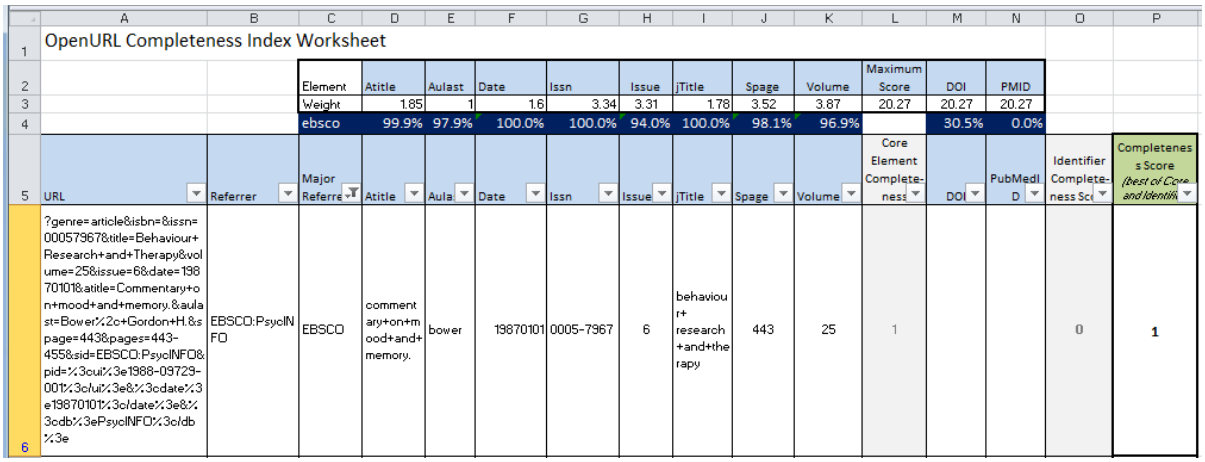
Or, if an organization has a reported problem with links coming from a certain OpenURL source, the Completeness Index can provide a quick way to diagnose if the problem is related to the data elements and if so, which ones. Then steps can be taken with the content provider to improve quality (see 4.1.3).

4.1.3 Analyzing the Completeness Index for a Problem

The IOTA website offers a series of analysis reports to investigate the OpenURLs from a given content provider to drill into fields provided.

When a lower than desired Completeness Index score is encountered, open the file containing the OpenURLs used to calculate the Completeness Index and use Excel to analyze further. Filter the OpenURLs by the Major provider in question, then sort the list of OpenURLs by Completeness Score (Smallest to Largest). The OpenURLs that appear first are the ones missing the most key data elements. Scan this list looking for patterns of missing fields.

To take a somewhat more scientific approach to the field analysis, add a simple control to the top of the OpenURL worksheet that will calculate the percentage of rows from a given content provider that contain data. Figure 8 shows field percentages added to row 4 of the spreadsheet.



OpenURL Completeness Index Worksheet

	Element	Atitle	Aulast	Date	Issn	Issue	jTitle	Spage	Volume	Maximum Score	DOI	PMID
3	Weight	1.85	1	1.6	3.34	3.31	1.78	3.52	3.87	20.27	20.27	20.27
4	ebsco	99.9%	97.9%	100.0%	100.0%	94.0%	100.0%	98.1%	96.9%		30.5%	0.0%

5	URL	Referrer	Major Referre	Atitle	Aula	Date	Issn	Issue	jTitle	Spage	Volume	Core Element Complete-ness	DOI	PubMedI D	Identifier Complete-ness Sc	Completenes s Score (best of Core and Identifi
6	?genre=article&isbn=&issn=00057967&title=Behaviour+Research+and+Therapy&volume=25&issue=6&date=19870101&atitle=Commentary+on+mood+and+memory.&aulast=Bower%2c+Gordon+H.&spage=443&pages=443-455&sid=EBSCO:PsycINFO&pid=%3cui%3e1988-09729-001%3c/ui%3e&%3cdate%3e19870101%3c/date%3e&%3cdb%3ePsycINFO%3c/db%3e	EBSCO:PsycINFO	EBSCO	commentary+on+mood+and+memory.	bower	19870101	0005-7967	6	behaviour+research+and+therapy	443	25	1			0	1

Figure 8: Example of OpenURL element frequency analysis by Referrer

Adding the field population percentages requires the use of the COUNTIF and COUNTIFS functions in Excel. In Figure 8, cell C4 contains the name of the Major Referrer to be analyzes. Cell D4 contains the following formula:

```
=(COUNTIF($C6:$C15004,$C$4) -
COUNTIFS($C6:$C15004,$C$4,D6:D15004,""))/COUNTIF($C6:$C15004,$C$4)
```

NOTE: In this example spreadsheet the OpenURLs can be found on rows 6 through 15,004; adjust these ranges in the formula to match your own spreadsheet.

The above formula counts the number of rows with a Major Referrer that matches what is typed in cell C4; it subtracts the number of rows from that Major Referrer with no data for the field in column "D", then divides that by the total rows from the provider. Format the result as a percentage.

Copy the formula in D4 and paste into Cells E4 through N4. Now you can see what percentage of a content provider's OpenURLs contain the field in question.

4.1.4 When a Content Provider is Frequently Missing Core Data Elements

The first thing to check when the analysis shows a content provider is frequently missing core data elements is whether or not the institution controls the OpenURL linking from that content provider. If so, it might be possible that there is a configuration problem that the institution can fix without involving the content provider. For example, a field may have been missed from a link template, or an element name might have been misspelled.

If the institution doesn't control the configuration of the content provider's OpenURL link, then the provider's customer support will need to be contacted. Inform them that some of the core data elements missing are missing from their URLS and request the problem be investigated and fixed. If possible, include the OpenURLs you used in the investigation.

4.2 Additional Items to Check

If missing data elements do not seem to be the cause of OpenURLs failing to generate full text links, here are some additional areas to check.

- **Check your holdings in the link resolver knowledge base.**
  OpenURLs may fail to generate full text links for other reasons than missing data elements. For example, an institution may not subscribe to the journal containing that article and thus the link resolver will omit the full text link because it failed the holdings check. Or if the organization does subscribe to the content, the holdings may not be correctly reflected.

- **Check the target link configuration.**
  If the organization has control over the link resolver's rules or templates used to configure the links to full text targets, check to make sure the settings for required data elements are correct. For example, are the correct elements listed as required elements—and are they spelled correctly? If the full text provider has both books and journal articles make sure the link rules are not requiring both an ISSN and an ISBN appear. (Two separate target links may be needed.)

- **Provide alternate links when full text is not available**.
  No one library has access to all published full text; therefore, the size and nature of the subscribed collection will determine the amount of full text end users will be able to link to directly. As mentioned above, an OpenURL with all the required data elements will not provide a link to the full text if that article is not included in the collection. When no subscribed full text is available, offer end users the option to request the article through your interlibrary loan department, or even search for it on Google.

- **Control end user expectations**.
  Many end users will automatically consider a link to an ILL form or the library catalog to be a "dead link"; the normal expectation is every click should result in immediate access to the full text. On some discovery and search systems, it may be possible to restrict the OpenURL links so that they only show if the search result is determined to be in the organization's collection. Implementing this is a good way of controlling when the links show. In addition to this, a topic on article linking as part of bibliographic instruction may be useful. If the link resolver allows, a note could also be added to the link menu when the ILL option appears explaining why the use is seeing this option.

Appendix A
A Method for Checking Link Resolution for Purposes of Calculation of the IOTA Completeness Index

The batch process used for checking link resolution for purposes of calculation of the IOTA Completeness Index in the Serials Solutions' 360Link link resolver is a modification of the standard quality assurance scripts used for end-to-end testing of mediated link production. Utilizing freely available open source software, it is a simple working example of similar scripts that can be implemented by vendors or libraries to do their own testing.

Written in Ruby 1.9.2, using Selenium Webdriver to interface with Firefox, the system might not be optimized for speed, but it does mean that each query is sent utilizing the user agent of the browser, so it appears, for all intents and purposes, identical to a live end user query. According to rules of good netiquette, a one-second delay should be built into the loop, to keep the queries from overwhelming the target server, though the loading delay of the browser is generally sufficient.

The script takes an OpenURL referring query as input and appends the proper Serials Solutions client identifier, in this case the 360Link QA demo library, and then sends the query to 360Link to look for knowledge base and article resolution. It parses the 360Link results page using regular expressions, looking for any article-level link on the page. If any such link exists, the referring URL is counted as a success for the purposes of the IOTA Completeness Index.

This simple success/failure criterion is then compared to the metadata fields in the initial referring query in order to produce a quantitative picture of link production.

Bibliography

ANSI/NISO Z39.84-2005 (R2010), *Syntax for the Digital Object Identifier*. Bethesda, MD: National Information Standards Organization, September 30, 2005. Available at: http://www.niso.org/standards/z39-84-2005/

ANSI/NISO Z39.88-2004 (R2010), *The OpenURL Framework for Context-Sensitive Services*. Bethesda, MD: National Information Standards Organization, April 15, 2005. Available at: http://www.niso.org/standards/z39-88-2004/

Chandler, Adam. *Results of L'Année Philologique Online OpenURL Quality Investigation, Mellon Planning Grant Final Report*. Ithaca, NY: Cornell University, February 2009. Available at: http://ecommons.library.cornell.edu/handle/1813/23171

Chandler, Adam, Glen Wiley, and Jim LeBlanc. "Towards Transparent and Scalable OpenURL Quality Metrics." *D-Lib Magazine*, March/April 2011, 17(3/4). doi: http://dx.doi.org/10.1045/march2011-chandler. Available at: http://dlib.org/dlib/march11/chandler/03chandler.html

Culling, James. *Link Resolvers and the Serials Supply Chain: Final Project Report for UKSG*. Oxford: Scholarly Information Strategies, May 21, 2007. Available at: http://www.uksg.org/projects/linkfinal

Glasser, Sarah. "Broken Links and Failed Access: How KBART, IOTA, and PIE-J Can Help." *Library Resources & Technical Services*, January 2012, 56(1): 14–23. Available at: http://www.questia.com/library/1G1-277520979/broken-links-and-failed-access-how-kbart-iota-and

ISO 26324:2012, *Information and documentation – Digital object identifier system*. Geneva, CH: International Organization for Standardization, April 23, 2012. Available at: http://www.iso.org/iso/home/store/catalogue_tc/catalogue_detail.htm?csnumber=43506

KBART (Knowledge Base And Related Tools) [website]. A joint NISO/UKSG project. Available at: http://www.niso.org/workrooms/kbart

Ponsford, Bennett Claire, Jane Stephens, and Robin R. Sewell. "Improving OpenURL Menus: User Testing of Revisions to SFX® Menus." *Serials Review*, September 2011, 37(3): 162–170. doi: http://dx.doi.org/10.1016/j.serrev.2011.03.016

Trainor, Cindi, and Jason Price. "Rethinking Library Linking: Breathing New Life into OpenURL." *Library Technology Reports*, October 2010, 46(7). Available at: http://alatechsource.metapress.com/content/k38116t41000/?p=55c47eeef5b147b0a889841a92c1f3e9&pi=18

Trainor, Cindi, and Jason Price. "Digging into the Data: Exposing the Causes of Resolver Failure." In: *Rethinking Library Linking: Breathing New Life into OpenURL, Library Technology Reports*, October 2010, 46(7): 15–26. Available at: http://alatechsource.metapress.com/content/m466322w0472g385/?p=efe2eec62aec47eab75e5150662ab46e&pi=2

Van de Sompel, Herbert, and Oren Beit-Arie. "Open Linking in the Scholarly Information Environment Using the OpenURL Framework." *D-Lib Magazine*, March 2001, 7 (3). doi: http://dx.doi.org/10.1045/march2001-vandesompel. Available at: http://www.dlib.org/dlib/march01/vandesompel/03vandesompel.html

Wakimoto, Jina Choi, David S. Walker, and Katherine S. Dabbour. “The Myths and Realities of SFX in Academic Libraries.” *The Journal of Academic Librarianship,* March 2006, 32(2): 127–136. doi: http://dx.doi.org/10.1016/j.acalib.2005.12.008.