

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This recommended practice may be revised or withdrawn at any time.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

Copyright © 2013 by the National Information Standards Organization

All rights reserved under International and Pan-American Copyright Conventions. For noncommercial purposes only, this publication may be reproduced or transmitted in any form or by any means without prior permission in writing from the publisher, provided it is reproduced accurately, the source of the material is identified, and the NISO copyright status is acknowledged. All inquiries regarding translations into other languages or commercial reproduction or distribution should be addressed to:
NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211.

ISBN: 978-1-937522-18-6

Contents

Foreword	v
Section 1: Introduction	1
1.1 Purpose and Scope	1
1.2 Terms and Definitions	1
Section 2: Determining Optimal Element Weights	5
2.1 The Stepwise Regression Approach	5
2.2 Preparing the Link Resolver Environment	7
2.2.1 Automated Testing Using a Batch Process	7
2.2.2 Checking for Full Text Targets	7
2.3 Selecting OpenURLs for the Stepwise Regression	8
2.4 Conducting the Tests	8
2.5 Calculating Optimal Element Weights	10
Section 3: Generating the Completeness Index	12
3.1 Reviewing the Theory Behind the Completeness Index	12
3.2 Selecting OpenURLs to Include in the Calculation	12
3.2.1 Source	12
3.2.2 Genre	12
3.2.3 Quantity	12
3.2.4 Time Period	12
3.3 Preparing for the test	12
3.4 Calculate Completeness Scores	15
3.5 Generate Completeness Index	15
Section 4: Analyzing Results and Affecting Change	17
4.1 Interpreting Completeness Index	17
4.1.1 Meaning of the Completeness Index Value	17
4.1.2 Using the Completeness Index	17
4.1.3 Analyzing the Completeness Index for a Problem	17
4.1.4 When a Content Provider is Frequently Missing Core Data Elements	18
4.2 Additional Items to Check	18
Appendix A A Method for Checking Link Resolution for Purposes of Calculation of the IOTA Completeness Index	20
Bibliography	21

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

Wherever used in this recommended practice, all terms that are trademarks or service marks are and remain the property of their respective owners.

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.

<u>Term</u>	<u>Definition</u>
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 referrer 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 .

<u>Term</u>	<u>Definition</u>
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. <i>See also success.</i> 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 **targets** 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 (**targets**), 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 **targets** 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 target**. 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.