



NISO-RP-9-2010



KBART: Knowledge Bases and Related Tools

*A Recommended Practice of the
National Information Standards Organization (NISO) and
UKSG*

Prepared by the
NISO/UKSG KBART Working Group

January 2010

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This recommended practice may be revised or withdrawn at any time.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)
One North Charles Street, Suite 1905
Baltimore, MD 21201
www.niso.org

Copyright © 2010 by the National Information Standards Organization and the UKSG. 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/UKSG copyright status is acknowledged. All inquiries regarding translations into other languages or commercial reproduction or distribution should be addressed to:
NISO, One North Charles Street, Suite 1905, Baltimore, MD 21201.

Printed in the United States of America and the United Kingdom
ISBN (13): 978-1-880124-83-3

KBART Phase I Recommended Practice

Table of Contents

Foreword	iii
1 Summary of Recommendations	1
2 Essential Terminology	1
3 Overview of OpenURL and Knowledge Bases	3
3.1 OpenURL Recap – Why? How?	3
3.2 Why Knowledge Bases Matter	4
3.3 The OpenURL Supply Chain: Roles and Benefits	4
3.3.1 Content Providers	4
3.3.2 Link Resolver Suppliers	7
3.3.3 Libraries	7
4 Overview of Problems in the Knowledge Base Supply Chain	8
4.1 KBART Scope	8
4.2 Illustration of Specific Data Accuracy Problems	9
4.2.1 Identifier Inconsistencies	9
4.2.2 Title Inconsistencies	10
4.2.3 Incorrect Date Coverage	10
4.2.4 Inconsistent Date Formatting	10
4.2.5 Inconsistencies in Content Coverage Description	10
4.2.6 Embargo Inconsistencies	11
4.2.7 Data Format and Exchange	11
4.2.8 Outdated Holdings Data	11
4.2.9 Lack of Customization	11
5 Guidelines for Effective Exchange of Metadata with Knowledge Bases	12
5.1 Transition to KBART	12
5.2 Exchange	13
5.2.1 Method of Exchange	13
5.2.2 Frequency of Exchange	13
5.2.3 Data Contacts	13
5.3 Data	13
5.3.1 Data Format	13
5.3.2 Data Fields	14
5.3.3 Error Reporting	19

KBART Phase I Recommended Practice

6 Education	20
6.1 Web Hub	20
6.1.1 Best Practice Guidelines	20
6.1.2 FAQ	20
6.1.3 Glossary	20
6.1.4 Quick Guides / Fact Sheets	20
6.1.5 Video demonstrations	20
6.1.6 Case studies	20
6.2 Outreach	21
6.3 FAQ Examples	21
6.3.1 General Questions	21
6.3.2 Content Provider/Publisher Questions	21
6.3.3 Librarian Questions	22
6.4 Guidelines on Improved Linking	22
7 Next Steps /Phase II	23
7.1 Recommendations and Further Discussion	24
7.1.1 Differences Between Aggregations and Individual Content Providers	24
7.1.2 Consortial Package Challenges	24
7.1.3 Institution-specific Metadata Transfer	24
7.1.4 Non-text Content Metadata Transfer	25
7.1.5 Review of Metadata Transfer for E-Books	25
7.1.6 Exchange of ERM Data	25
7.1.7 Compliance with KBART Recommendations	25
Appendix A : Data Exchange Samples	26
Glossary of Terms	29
Bibliography	33

Figures

Figure 1: User Journey via OpenURL	3
Figure 2: Flow of data through the supply chain	6

KBART Phase I Recommended Practice

Foreword

KBART Working Group

In 2007, UKSG, a nonprofit organization that connects the information community, commissioned a report, *Link Resolvers and the Serials Supply Chain*¹. This report identified and described a range of problems affecting the efficiency of OpenURL linking. The report recommended (in section 7.1.1) the creation of a group that would determine and promote “best practice” solutions for the overall community to improve the exchange of metadata with knowledge bases.

In conjunction with the National Information Standards Organization (NISO), UKSG proceeded to set up a working group that would bring together members of all parts of the electronic resources supply chain to address the problems identified in the UKSG report, create guidelines, and propose solutions. The joint NISO/UKSG KBART (Knowledge Bases And Related Tools) Working Group was established in December 2007 and this Recommended Practice is the result of its initial phase. Information about the group’s processes and membership is given in the Foreword.

The formation of the KBART Working Group was publicized widely by UKSG and NISO, and representatives from the OpenURL supply chain were invited to express their interest. Approximately 50 expressions of interest were received. Co-chairs, appointed by the UKSG and NISO leadership committees, selected 12 core Working Group members to represent equally the different stakeholders in the supply chain. Members are listed below. Others who had expressed interest were invited to join the monitoring Interest Group, which received regular reports on the group’s progress and was asked to help with reviewing the Recommended Practice prior to its publication.

The group met monthly by conference call between December 2007 and December 2009. Members were divided into sub-groups and allocated specific areas of the report to work on. Progress was then reported back to the group each month for discussion and prioritization of ongoing activities.

Scope and charge

The NISO/UKSG KBART Working Group’s scope focuses on problems in the information supply chain that relate to the data supplied to knowledge bases. This specifically excludes wider problems with OpenURL linking, which fall either within the remit of OCLC, the Maintenance Agency for the OpenURL standard (ANSI/NISO Z39.88-2004, *The OpenURL Framework for Context-Sensitive Services*), or within other NISO working groups. The group has also focused specifically on data relating to content holdings rather than on bibliographic data about individual titles, which does not need to be updated as regularly as holdings data.

The KBART Working Group’s charge is to improve the supply of data to link resolvers and knowledge bases, in order to improve the efficiency and effectiveness of OpenURL linking. This is to be achieved by providing best practice guidelines, educational materials and events, and a web hub to act as a central resource for knowledge base information.

¹ Culling, J. *Link Resolvers and the Serials Supply Chain*. Oxford: Scholarly Information Strategies, 2007. Available at <http://www.uksg.org/projects/linkfinal>

KBART Phase I Recommended Practice

NISO Topic Committee Members

The NISO Discovery to Delivery Topic Committee had the following members at the time it approved this Recommended Practice:

Susan Campbell
College Center for Library Automation

Larry Dixson
Library of Congress

David Fiander
University of Western Ontario

Mary Jackson
Auto-Graphics

John Law
ProQuest

Tony O'Brien (Co-Chair)
OCLC

Norman Paskin
Tertius Ltd

Matt Turner
Mark Logic Corporation

Tim Shearer (Co-Chair)
University of North Carolina

UKSG Main Committee Members

UKSG's Main Committee had the following members at the time it approved this Recommended Practice:

Beverley Acreman
Taylor & Francis

Jo Connolly
Swets

Lesley Crawshaw
University of Hertfordshire

Richard Gedye
Oxford University Press

Claire Grace
The Open University

Ian Hames
ebrary, inc

Helen Henderson (Editor, *Serials*)
Ringgold Ltd

Tony Kidd (Chair)
University of Glasgow

Ross MacIntyre
Mimas, The University of Manchester

Sarah Pearson
University of Birmingham

Ed Pentz (Treasurer)
CrossRef

Kate Price (Education Officer)
University of Surrey

Charlie Rapple (Marketing Officer)
TBI Communications

Graham Stone (Secretary)
University of Huddersfield

Jill Taylor-Roe
University of Newcastle

Diane Thomas
Gale Cengage Learning EMEA

Hazel Woodward (Editor, *Serials*)
Cranfield University

KBART Phase I Recommended Practice

NISO/UKSG KBART Working Group Members

The following individuals served on the NISO/UKSG KBART Working Group, which developed and approved this Recommended Practice:

Phil Caisley

BMJ Group

Adam Chandler

Cornell University Library

Anna Gasking

Informa plc

Simon Haggis

BMJ Group

Nettie Lagace

Ex Libris Group

Peter McCracken (Co-Chair,
representing NISO)

Serials Solutions

Chrissie Noonan

Pacific Northwest National Laboratory

Sarah Pearson

University of Birmingham

Oliver Pesch

EBSCO Industries Inc

Jason Price

Claremont Colleges / Statewide
California Electronic Library Consortium

Charlie Rapple (Co-Chair, representing
UKSG)

TBI Communications

Elizabeth Stevenson

University of Edinburgh

Margery Tibbetts

California Digital Library

Thomas P. Ventimiglia

Princeton University

Jenny Walker

Consultant

1 Summary of Recommendations

A link resolver is a tool that helps library users connect to their institutions' electronic resources. The data that drives such a tool is stored in a knowledge base.

The quality of a knowledge base depends heavily on data that content providers (publishers, aggregators, etc.) send to the knowledge base developer. Errors in this data often propagate to the knowledge base. Furthermore, because there is no standard format for such data, knowledge base developers must expend much effort converting title lists from different providers to a single format, which may introduce additional errors or make error-checking difficult.

The NISO/UKSG KBART Recommended Practice recommends some best practices for formatting and distributing title lists. By making some small adjustments to the format of their title lists, content providers can greatly increase the accessibility of their products. These recommendations are designed to be intuitive, easy for content providers to implement, and easy for knowledge base developers to process.

- Section 2 provides the essential terminology needed to understand the Recommended Practice; a full glossary is included at the end of this document.
- Section 3 gives a brief overview of OpenURL link resolving, knowledge bases, and the information supply chain around them.
- Section 4 identifies some typical problems with knowledge bases, their causes, and their impact on the user experience.
- **Section 5 describes in detail a set of solutions and best practices that will help avoid these problems.**
- Section 6 explains the role KBART plans to take in supporting the adoption of these practices.
- Section 7 proposes some ways in which KBART's work may be expanded in the future.

2 Essential Terminology

A comprehensive glossary to help readers understand the terminology used in this report is given at the end of this report. A subset of critical terms and their definitions is given here since they will be used extensively throughout the following pages, and understanding these terms within the context of KBART is key to understanding the report.

Note that Phase I of KBART's work, resulting in this Recommended Practice, has focused on the supply chain for text-based materials such as journals and e-books, as this is the area where OpenURL linking is already prevalent and therefore where the majority of problems have occurred to date. The terms selected and the definitions provided should be understood within this context. See Section 7 for more information on potential work planned for the next phase of work.

KBART Phase I Recommended Practice

Appropriate copy

One or more versions, among many, that are most appropriate for a specific user in a specific situation at a given institution. This is likely to be a version of which they are entitled to access the full text, probably because of a subscription paid for by the library.

Content provider

A vendor—generally a publisher, aggregator, or full-text host—that offers content for sale or lease to libraries. This may also include abstracting and indexing services, subscription agent gateways, and other sources of OpenURL links.

Inbound linking (syntax)

Links into a website from other online resources. A content provider is enabling inbound linking if they make publicly available a link-to syntax enabling others to predict the URL of pages within their website, at various levels (e.g., journal home pages, tables of contents, or specific articles).

Knowledge base

An extensive database maintained by a link resolver vendor, containing information about electronic resources such as title lists, coverage dates, inbound linking syntax, etc. The knowledge bases can be customized by individual institutions to reflect their local collections—for example, which titles can be accessed electronically and which resources are owned by the library in print format. This is typically referred to as the local knowledge base. (This report will use the two-word phrase “knowledge base,” but “knowledgebase” is also commonly used.)

Link resolver

A “link resolver,” or “link server,” is a software tool that deconstructs an OpenURL, separates out the elements that describe the required article, and uses these to create a predictable link to the appropriate service(s) identified by the user’s library.

Link-to syntax

The formula by which links to specific pages within a website can be constructed, usually consisting of a base URL and a string of metadata / identifiers. Some content providers follow the OpenURL syntax to enable inbound linking; others base their link-to syntax on proprietary, but predictable, identifiers.

OpenURL

The OpenURL standard (ANSI/NISO Z39.88-2004, *The OpenURL Framework for Context-Sensitive Services*) specifies the syntax for transporting metadata from information resources (sources) to an institutional link resolver and thence to library services (targets).

Source

The resource that creates an OpenURL and thereby links to a link resolver. The source can be understood as the overall website (e.g., database, publisher platform etc.) or as a specific citation within it.

Target

The resource that is linked to by a link resolver. Example targets include content in publisher platforms, institutional catalogues, or repositories and content gateways.

3 Overview of OpenURL and Knowledge Bases

Version 0.1 of the OpenURL was introduced in 1999; version 1.0 became a NISO standard (ANSI/NISO Z39.88-2004, *The OpenURL Framework for Context-Sensitive Services*) in 2004. It has been adopted throughout the scholarly information supply chain to support improved linking between resources. A range of suppliers has developed tools that support effective OpenURL implementation, and a link resolver (whether licensed or homegrown) has become a key part of any research library's toolkit.

3.1 OpenURL Recap – Why? How?

Conventional reference linking initially involved hard-coding links between one content provider and another. As a result, users were often linked to the “wrong” version of an article, i.e., one that they were not licensed to access. In the worst case scenario, this would result in a user undertaking a document delivery or pay-per-view transaction to obtain an article that might actually have been licensed elsewhere by their library. This is known as the “appropriate copy” problem.

The OpenURL was developed to perform “context-sensitive” linking, whereby links are flexible and able to take into account the user's institutional affiliations and the licenses of that institution. Following ratification as an ANSI/NISO standard, OpenURL linking has been widely adopted. A basic user journey via OpenURL is illustrated in Figure 1.

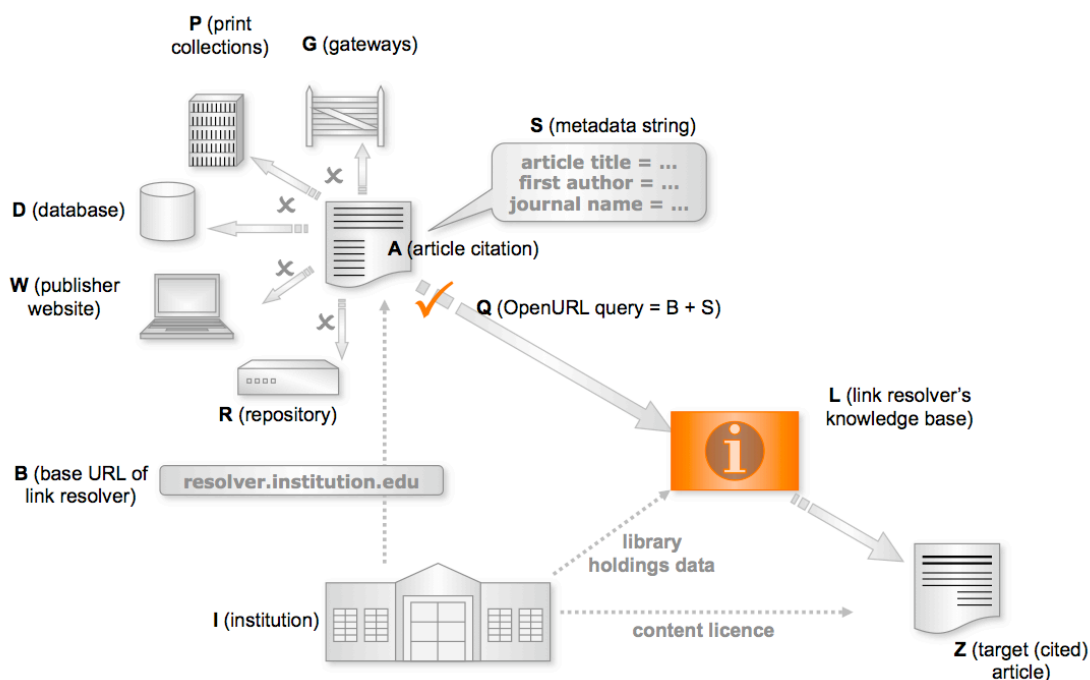


Figure 1: User Journey via OpenURL

The OpenURL uses a link resolver (L) to transport a user from A (a citation) to Z (a copy of the cited document which is licensed by the current user), by way of an OpenURL query (Q), which appends a string (S) of metadata about the cited article to the base URL (B) of the current user's institution (I). This is a more effective alternative than hard-coded links to other resources, such as a subscription agent gateway (G), the library's print holdings (P), aggregated databases (D), publishers' own websites (W), or repositories (R).

3.2 Why Knowledge Bases Matter

Knowledge bases are key to the process of OpenURL linking because they not only know where content is, but they also know which versions of specific objects a particular institution's users are entitled to access. Knowledge bases are the only means by which users can be sure to reach an "appropriate copy."

If data provided to knowledge bases is incomplete, inaccurate, out of date, or in some other way "bad," the efficacy of the OpenURL standard is undermined such that it can often become useless. As a result, the NISO/UKSG KBART Working Group was formed to analyze the problems within the supply chain and create guidelines to resolve the most common or high-impact problems.

While the focus of KBART is on data exchange among and between knowledge bases, it is acknowledged that the inclusion and correct encoding of data within OpenURLs is equally critical to the success of OpenURL linking.

3.3 The OpenURL Supply Chain: Roles and Benefits

The OpenURL supply chain includes many stakeholders with many connections in between them. Each stakeholder has specific responsibilities with regard to the data they share with other members and stakeholders. The data transfer responsibilities of parties within the OpenURL supply chain is mapped out in Figure 2.

In an attempt to identify what is expected of each participant of the supply chain, and to identify areas where confusion or inaccuracy can be introduced, we attempt to describe the expected responsibilities for each stakeholder here.

At the basic level, the following stakeholders in the OpenURL supply chain—content providers, link resolver providers, and libraries—are required to carry out the following tasks, and in an efficient supply chain, would reap the following benefits.

3.3.1 Content Providers

Content providers can include publishers, online delivery providers, subscription agent gateways, full text aggregators, and others. But most metadata, like most content, originates from a publisher. In many cases, the metadata that is transferred in subsequent steps begins with the publisher, so if it is incorrect at the start, it will remain incorrect for most, if not all, of the remainder of the supply chain.

Role: Content providers can be both a source of and a target for OpenURL links. Adherence to the OpenURL standard requires a content provider to be able to create compliant OpenURLs from their citations. For other OpenURL sources to be able to create links to the content provider, that provider must also make available accurate metadata about its holdings. Currently there is no standard format for such data; part of KBART's mission is to create best practice guidelines in this area.

Functions:

- Deliver articles—with appropriate metadata—for publication or hosting
- Receive full-text content from publishers
- Create metadata at full text or abstract level
- Host full text and provide related functionality on behalf of publishers

KBART Phase I Recommended Practice

- Create OpenURL links
- Send holdings lists to knowledge base developers
- Send holdings lists to libraries (for their unique holdings)
- Provide MARC records to libraries

Benefits: Timely transmission of accurate holdings metadata to link resolver suppliers benefits content providers by creating a smoother user experience—thereby reducing the cost of customer service and improving the publisher’s reputation—and driving more traffic to their content. Increased traffic supports the publisher, editor, and author’s objective to ensure maximum visibility, usage, and reach of their content. It can also support usage-based revenue streams, and is a key factor in purchasing decisions when libraries come to renew or cancel content licensing agreements.

Key to knowledge base success:

- Ensuring that accurate holdings data is provided to link resolver owners on a regular schedule and that known errors are corrected as quickly as possible.