



**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This recommended practice may be revised or withdrawn at any time.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](http://www.niso.org)).

### **Published by**

National Information Standards Organization (NISO)  
One North Charles Street, Suite 1905  
Baltimore, MD 21201  
[www.niso.org](http://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

## Table of Contents

<b>Foreword</b>	<b>iii</b>
<b>1 Summary of Recommendations</b>	<b>1</b>
<b>2 Essential Terminology</b>	<b>1</b>
<b>3 Overview of OpenURL and Knowledge Bases</b>	<b>3</b>
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
<b>4 Overview of Problems in the Knowledge Base Supply Chain</b>	<b>8</b>
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
<b>5 Guidelines for Effective Exchange of Metadata with Knowledge Bases</b>	<b>12</b>
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

<b>6 Education</b>	<b>20</b>
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
<b>7 Next Steps /Phase II</b>	<b>23</b>
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
<b>Appendix A : Data Exchange Samples</b>	<b>26</b>
<b>Glossary of Terms</b>	<b>29</b>
<b>Bibliography</b>	<b>33</b>

## Figures

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

## 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*<sup>1</sup>. 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.

---

<sup>1</sup> 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.

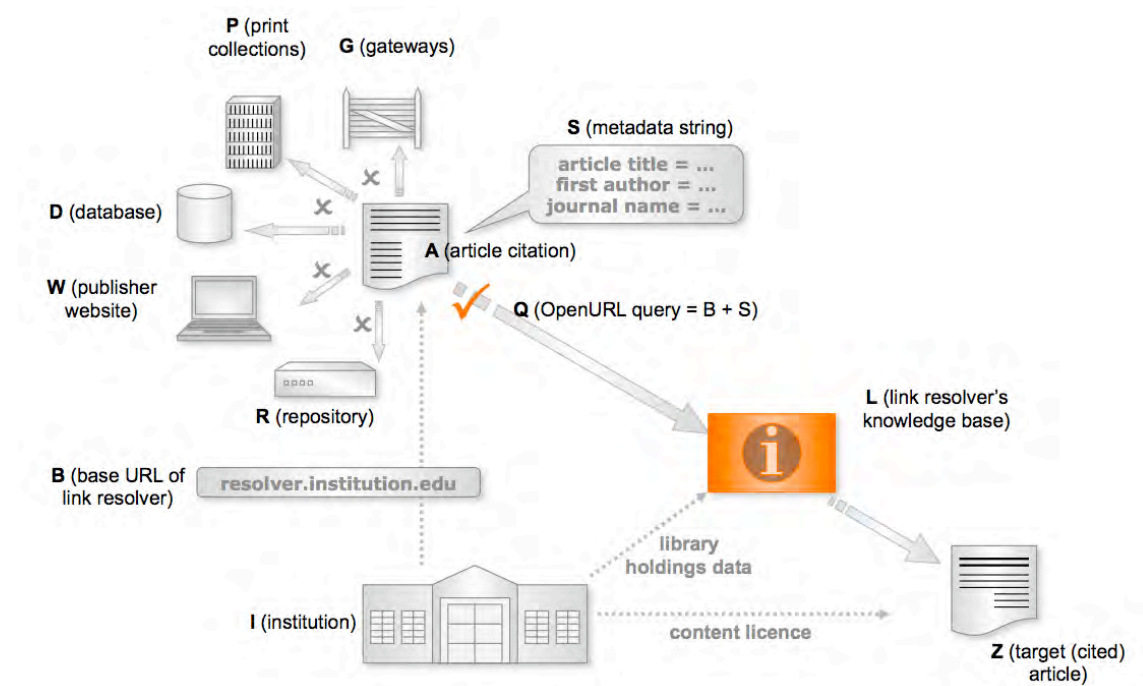


Figure 1: User Journey via OpenURL

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

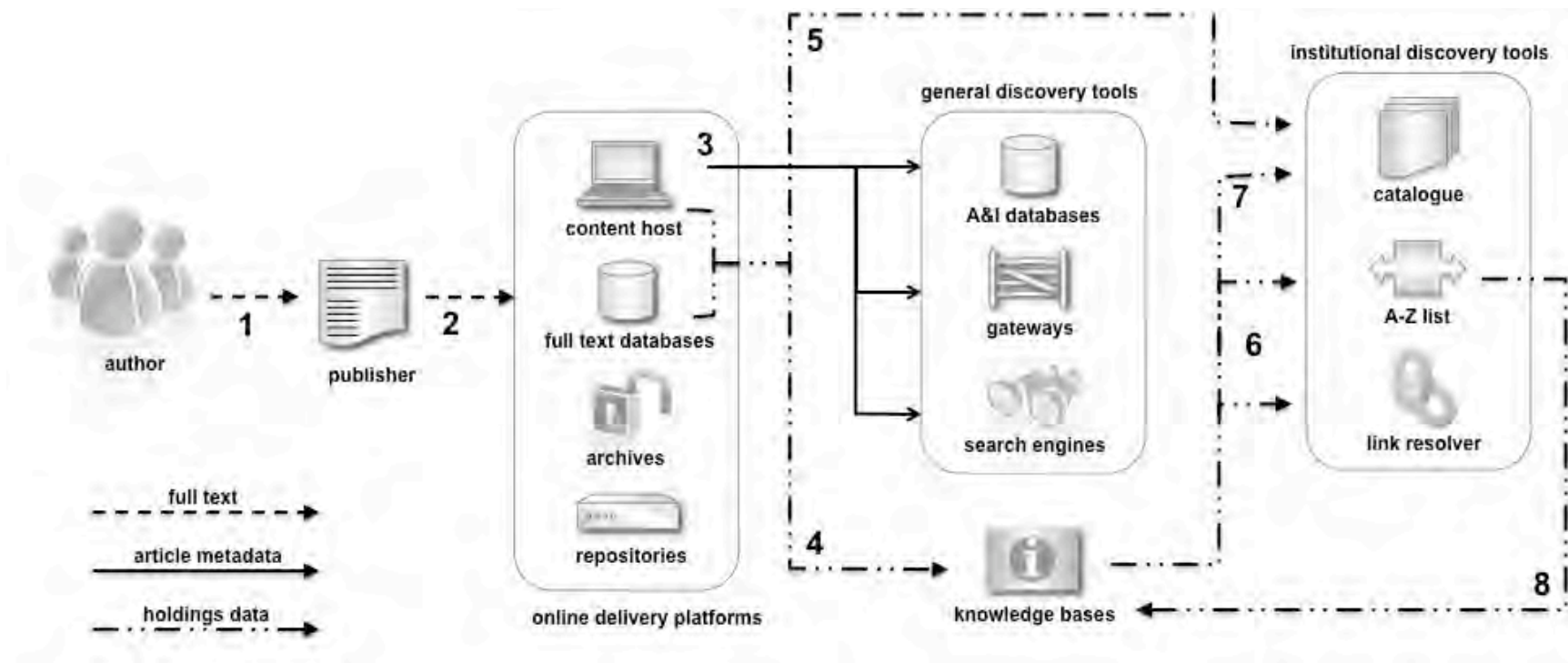
- Create OpenURL links
- Send holdings lists to knowledge base developers
- Send holdings lists to libraries (for their unique holdings)
- Provide MARC records to libraries

**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.

## KBART Phase I Recommended Practice



**Figure 2: Flow of data through the supply chain**

The author supplies manuscripts to the publisher (1). The publisher sends final article full text and associated metadata to one or more online delivery services (content hosts, aggregated full text databases, archive sites, etc.) (2). The content host is frequently responsible for delivering the article metadata to other discovery tools, such as A&I (abstracting and indexing) databases, subscription agent gateways, and search engines (3). Content hosts and full-text databases also deliver holdings information and sometimes MARC records to knowledge base developers (4) or directly to the institution (5) in order to inform those systems about the extent of content on the delivery platform. Some knowledge base services will then send holdings data to institutional A-to-Z lists and link resolvers (6); the institution uses this information to configure its own collection. Knowledge base developers may also send MARC records to the institution for inclusion in the institutional OPAC (7). In this scenario, the institution must supply its knowledge base developer with details of its holdings for content such as e-journals, etc. (8).

## KBART Phase I Recommended Practice

### 3.3.2 Link Resolver Suppliers

**Role:** Link resolvers are configured to receive OpenURL links from content provider sources. They extract metadata about the target article (or other object) from the OpenURL, then compare this information to the knowledge base, which contains data about all the content licensed by the link resolver's owner. The knowledge base indicates *whether* the article is available to individuals associated with the license-holding institution; if so, *where* it is hosted; and *how* to connect the user to it. The link resolver then puts together a link to the target article. This process is largely transparent to the user. The link resolver supplier maintains the technology of the link resolver, the data within the global knowledge base and creates or assists with localization of knowledge bases.

**Functions:**

- Regularly receive current holdings lists from content providers for normalization, processing, and inclusion in the global knowledge base
- Receive details of content providers' inbound linking syntaxes to accompany collection data for link targets
- Provide means for content providers to test their OpenURLs and metadata supply
- Make content providers aware of their link resolver base URLs
- Check (and possibly document) the level of OpenURL support
- Generate and distribute MARC records for library's holdings

**Benefits:** Timely transmission of accurate metadata to link resolvers, and good communication between link resolver supplier, library and content provider, enables the link resolver supplier to provide a more accurate and current service, and thereby to fulfill their goal of connecting users to licensed content. Timely transmission of accurate holdings metadata will also reduce any costs associated with checking, cleaning, and maintaining knowledge base holdings data.

**Keys to knowledge base success:**

- Ensuring that knowledge base is regularly updated and incorporates corrections provided by other members of the supply chain.

### 3.3.3 Libraries

As the final presenter of data to the end user, the library has an important role in ensuring that the presentation of metadata is accurate and useable. In addition, it is the library and the end user who are most likely to notice errors in OpenURL data or in citation data, and are therefore best placed to notify other members of the electronic resources supply chain about those errors.

**Role:** Libraries build, purchase or license link resolvers in order to maximize efficient access to all the electronic resources they license. A library registers key identifying information about its link resolver (such as the base URL) with those content providers that it wishes to act as a source of OpenURL links. It is likely that each library will need to customize the knowledge base consulted by its link resolver in order to reflect its own print and online holdings.

**Benefits:** Effective link resolver deployment benefits libraries by maximizing the usage

## KBART Phase I Recommended Practice

(and therefore the return on investment) of the content they license, and improves the experience and success rate of their users as they navigate the research network. This can reduce reference and e-resource support queries. Timely transmission of accurate holdings metadata will also reduce the current costs and burden of checking, cleaning, and maintaining knowledge base holdings data.

### Functions:

- Receive and apply knowledge base updates on a regular schedule
- Register the library's link resolver details accurately with each content provider
- Where the link resolver is hosted by the knowledge base developer, supply collection holdings data to knowledge base developer to enable customization of the knowledge base
- Activate link targets in the link resolver supplier's knowledge base. The library must collect details of collections for updating the knowledge base. The updating is done manually using tools provided by the link resolver supplier and, where applicable, by customizing the holdings list provided by the Online Delivery Provider.
- Notify the link resolver supplier if the knowledge base data is inaccurate

### Keys to knowledge base success:

- Providing data to customize the knowledge base according to local holdings.

## **4 Overview of Problems in the Knowledge Base Supply Chain**

### **4.1 KBART Scope**

The goal of the OpenURL is to connect users with resources that they are licensed to access. However, some basic problems can prevent the technology from working effectively:

- Lack of uptake of OpenURL technology by some providers
- Poor metadata held in knowledge bases
- Inaccurate implementation of OpenURL syntax by OpenURL sources
- Poor inbound URL syntax management by OpenURL targets

It is within the scope of KBART's charge is to assess those problems that result from or relate to the data provided to knowledge bases, encompassing items 1 and 2 above. This Recommended Practice does not consider problems that result from or relate to maintenance of the OpenURL standard documentation, since these are within the responsibility of the OpenURL standard's Maintenance Agency, currently OCLC (<http://www.oclc.org/research/activities/openurl/default.htm>). OCLC facilitates the OpenURL community through an e-mail list ([openurl@oclc.org](mailto:openurl@oclc.org)) that facilitates discussion and serves as a forum in which the community can resolve issues. Details of the registry for the OpenURL framework are available at <http://alcme.oclc.org/openurl/servlet/OAIHandler?verb=Identify>.

## KBART Phase I Recommended Practice

However, alongside recommendations provided here relating to metadata transfer, content providers should also supply details of their linking syntax at all levels to knowledge base developers. This will ensure accurate linking to content at title, issue, and article level, which, along with accurate metadata, will ensure that end users can successfully access appropriate content.

The Working Group considered the following areas relating to poor metadata about OpenURL targets, discussed in Section 4.2 below, to be within scope for KBART:

- 4.2.1 Identifier Inconsistencies
- 4.2.2 Title Inconsistencies
- 4.2.3 Incorrect Date Coverage
- 4.2.4 Inconsistent Date Formatting
- 4.2.5 Inconsistencies in Content Coverage Description
- 4.2.6 Embargo Inconsistencies
- 4.2.7 Data Format and Exchange
- 4.2.8 Outdated Holdings Data
- 4.2.9 Lack of Customization

---

### 4.2 Illustration of Specific Data Accuracy Problems

In this section we provide examples of how users are impacted by poor data management within the knowledge base supply chain. We provide indicative solutions; full recommendations for best practice are given in Section 5, Guidelines for effective exchange of metadata with knowledge bases.

#### 4.2.1 Identifier Inconsistencies

**Core problem:** Re-use of the ISSN (for example, when a journal changes title) or ISBNs (for example, among multiple titles from the same publisher) causes confusion on the part of the link resolver, as the unique identifier is no longer unique.

**User impact:** The OpenURL syntax uses the unique ISSN as a key component. When a citation (i.e., source) uses an accurate ISSN, but a full-text resource (i.e., target) uses an inaccurate ISSN—such as one for a previous or subsequent title—the link resolver may be unable to resolve the difference, and fail to return accurate links to the content being sought.

**Solutions:** The accurate use of the ISSN is critical in ensuring successful resolution of OpenURL-based links. The content provider should ensure that the ISSN being supplied in the input file for the knowledge base is accurate in relation to the work described. To avoid inappropriate reuse of ISSN when assigning ISSN to new and changing works, publishers are encouraged to refer to local ISSN agency guidelines (e.g., <http://www.bl.uk/bibliographic/issn.html> or <http://www.loc.gov/issn/>) or to the international ISSN agency <http://www.issn.org>.

KBART also expects that knowledge base developers be capable of successfully managing multiple ISSN for a given title. Most titles that are available both in print and online have more than one valid ISSN assigned to them. This is a common occurrence, and a link resolver knowledge base must be able to manage this successfully.



## KBART Phase I Recommended Practice

### 4.2.2 Title Inconsistencies

**Core problem:** Variant titles, particularly caused by misspellings, title variations, or the use of previous or subsequent titles, lead to matching problems.

**User impact:** Most resolvers have a way to handle variations in capitalization and punctuation, but may not be able to resolve abbreviations (e.g., “NEJM” for the *New England Journal of Medicine*), regional variations in spelling (e.g., “labour” vs. “labor”), or numbers written out as words (e.g., “1st” vs. “first”). The link resolver may fail to find the correct title—or any title – when it is attempting to resolve an OpenURL query. The user will not reach the content he seeks.

**Solutions:** Best practice is for all systems to use a form of the title that appears as a main title in a standard cataloging source, such as the CONSER record set, the OCLC WorldCat database, or the ISSN Registry. If no such resource contains a record for the title, then the title should be reproduced as it appears on the cover of the print edition. In addition, it is expected that knowledge bases are able to manage variant titles within their system.

### 4.2.3 Incorrect Date Coverage

**Core problem:** The presentation of inaccurate dates by content providers, or failure to update dates in the knowledge base by knowledge base developers, causes inaccurate results or failed links in the link resolver.

**User impact:** Individuals may be told that a resource they seek is not available when it is (for example, if the file is not updated to reflect the most recent issues online) or that a resource is available when it is not (for example, if content has been removed but is still listed in a holdings list). In the former case, valid usage will decrease; in the latter, individuals will eventually shy away from using the resource if attempts to follow links to it regularly result in dead ends.

**Solutions:** Content providers should provide accurate holdings files of coverage dates in a timely fashion; link resolver owners should ensure they are processing holdings files correctly and promptly.

### 4.2.4 Inconsistent Date Formatting

**Core problem:** Without a standard format, coverage dates can be interpreted incorrectly. (For example, does 20080305 signify 5 March 2008 or 3 May 2008?) The existence of multiple dates for an article (for example, online posting date versus issue publication date, where these differ) can also prevent a match being made between a citation and the knowledge base.

**User impact:** Inaccurate date information prevents the knowledge base from matching a source citation to a target resource, which in turn prevents end users from accessing resources, causing frustration and reducing usage.

**Solutions:** The content provider should use the most current version of standard ISO 8601:2004 (*Data elements and interchange formats – Information interchange – Representation of dates and times*) date syntaxes when interpreting and displaying date ranges.

### 4.2.5 Inconsistencies in Content Coverage Description

**Core problem:** Some providers offer only limited portions of the print version’s content online, or publishers may only provide limited content to aggregators. As a result, whole articles within the stated holdings range may be missing. In other cases, aggregated

## KBART Phase I Recommended Practice

content is represented as complete at the title level and includes all text, but lacks tables, figures, notes, or some other integral part of the article.

**User impact:** Individuals fail to obtain some or all of the content they are expecting to find; in cases of incomplete articles, they may not realize they do not have all that is offered by the original content provider.

**Solutions:** Content providers should indicate what kind of content is being offered. A set of terms to describe coverage is given in Section 5.3.2.15.

### 4.2.6 Embargo Inconsistencies

**Core problem:** Knowledge base developers are often uncertain as to how to apply varying embargo terms, as the same terms can be used by different content providers in different ways.

**User impact:** Embargo terminology is complex, particularly in instances of a “1-year embargo,” which might mean the previous *12 months* are not available, or that the current (or previous) *calendar year* is not available (i.e., from January of a certain year to the present). Users cannot know with certainty if the article they seek will be found in a given database when the issue date falls in or near the range of the embargo period.

**Solutions:** Consistent usage of the ISO 8601 duration syntax, which allows multiple types of embargo to be described, can mitigate this problem. See Section 5.3.2.14 for details.

### 4.2.7 Data Format and Exchange

**Core problem:** Data errors or non-standard exchange of data can slow the incorporation of that data into the knowledge base. Data that does not conform to previously agreed-upon formats can also be a cause of failed links.

**User impact:** Inaccurate and outdated information leads to links that fail to resolve correctly, and prevents users from reaching content to which they should have access.

**Solutions:** Content providers should conform to the simple guidelines in this report and allocate a member of staff to monitor the smooth transition of data. Link resolver suppliers should provide content providers with test account access so that they can verify data on their publications as represented in the knowledge base.

### 4.2.8 Outdated Holdings Data

**Core problem:** Outdated holdings data leads to inaccurate presentation of data.

**User impact:** Data presented in knowledge bases is only as accurate as the data imported to them; if current information is not available, users are unable to reach recent content that they are entitled to access.

**Solutions:** Content providers should supply holdings information on a regular basis, dependent on the frequency of their publication schedule. Link resolver suppliers should enable users of library catalogues to see when data was last generated, and when it will next be updated. Link resolver suppliers should inform content providers and librarians of their schedule for updating knowledge bases and confirm when updates have been made.

### 4.2.9 Lack of Customization

**Core problem:** Holdings data generally shows the broadest range of dates available to all customers, rather than just the titles and dates available to a single institution.

## KBART Phase I Recommended Practice

**User impact:** Librarians must spend time customizing these lists to make them accurately reflect the holdings at their institutions. This provides a significant opportunity for the introduction of errors, which will then lead to individuals not accessing or using resources that a publisher offers.

**Solutions:** Link resolver suppliers and content providers should work together to exchange library-specific holdings files, which would allow individual institutions to better manage data relative to their specific institution. Given the critical importance as well as the complexity of the customization issue, KBART intends to investigate it further during a later phase. Even if the provider cannot generate a separate file for each institution, it could still generate specific holdings files for popular subscription plans or content packages.

## 5 Guidelines for Effective Exchange of Metadata with Knowledge Bases

---

Many content providers and knowledge base developers are already successfully exchanging metadata, and this report is not intended to detract from or interfere with such existing processes. However, it is evident that many others are unsure about how best to exchange metadata. Therefore, we propose entry-level guidelines and instructions to enable exchange of essential metadata.

Our recommendations are based on those methods and data fields that have proven to be effective or valuable in our combined experience. In many cases there are acceptable alternatives, but for clarity and simplicity we are distilling our experience into a single recommendation, where possible. Our recommendations also exclude information that is more appropriately classed as bibliographic data than holdings metadata, for example, language, alternate titles, content type, or relationship to other titles. Since this data is more static than holdings data and does not need to be exchanged on such a regular basis, it is outside of the scope of the KBART recommendations.

---

### 5.1 Transition to KBART

For those content providers that already provide metadata files to knowledge base developers (for example, in OAI or ONIX-SOH format), this set of recommendations can be used as a guide to review their current provision with a view to incorporating missing mandatory fields recommended in this document. It may be that current methods of file transfer and naming conventions adopted by these content providers is adequate and no change may be required to file transfer and naming.

For those content providers that don't send metadata files relating to their content already, these recommendations can be taken as a full set of implementation guidelines.

Content providers who are participants in CrossRef can contact CrossRef for assistance in some areas of metadata exchange; CrossRef already manages nearly all of the information defined below, and is exploring options for serving as a central hub to distribute its members' information to metadata users.

### 5.2 Exchange

#### 5.2.1 Method of Exchange

Content providers should post holdings data to a website or FTP site for download by link resolver suppliers. This minimizes the effort involved in the transaction for both parties.

FTP (File Transfer Protocol) is a simple protocol for allowing users to exchange files. It allows access to the metadata to be restricted to authorized users, though content providers should recognize that broad dissemination of information about accessing their content is in their best interest, and multiple link resolvers (including libraries managing their own link resolvers) should be able to access the data.

Posting to the web or to an FTP site is preferable to e-mail exchange because it is harder to incorporate e-mail into an automated process for checking, validating, and uploading new data. E-mail exchanges are also subject to length restrictions, spam filters, and individuals' availability. However, if posting to the web or FTP site is unachievable, then e-mail is an acceptable alternative. The data from the tab-delimited file (see Section 5.3.1.1) should be placed in the body of the e-mail; the e-mail's subject line should also follow the naming convention given in Section 5.3.1.2.

#### 5.2.2 Frequency of Exchange

A monthly metadata update is recommended; however, when content is added less frequently than monthly, content providers may then choose a less frequent schedule for updates. Alternatively, providers may update data more frequently than once a month if they wish.

#### 5.2.3 Data Contacts

Both the content provider and the knowledge base developer should designate specific staff members to be responsible for data files and exchange. Doing so expedites resolution of any problems that may develop. Content providers will need to inform the designated knowledge base developer contact about any changes to the data exchange process. Knowledge base developers will need to inform the designated content provider contact about any errors in the data. Both contacts will need to take responsibility for passing messages to the appropriate staff within their organization and ensuring appropriate action is taken. To facilitate this relationship, the KBART Working Group will also provide a web form that will be editable by the community to add details of content provider and knowledge base developer contacts. A link to this web form will be provided on the KBART web pages.

---

### 5.3 Data

#### 5.3.1 Data Format

**5.3.1.1** Content providers should provide metadata formatted as tab-delimited values. This is a generic format that minimizes the effort involved in receiving and loading the data, and reduces the likelihood of errors being introduced during exchange. Tab-delimited formats are preferable to comma-separated formats, as commas appear regularly within the distributed data and, though they can be "commented out," doing so leaves a greater opportunity for error than the use of a tab-delimited format. Tab-delimited formats can be easily exported from all

## KBART Phase I Recommended Practice

commonly used spreadsheet programs.

- 5.3.1.2 The file should be entitled “[ProviderName]\_AllTitles\_[YYYY-MM-DD].txt”. For example, `JSTOR_AllTitles_2008-12-01.txt`.
- 5.3.1.3 The provider name should be the web domain at which your data is hosted (but without the punctuation). For example, `jstor` or `ebscohost`. This ensures that your data is clearly distinguished from data provided by others with similar package names. Also, the file name should be consistent for each metadata file deposited.
- 5.3.1.4 Separate files should be produced for each package of content that the provider offers. Files should be named as customers would expect to see it labelled in the knowledge base, using the syntax “[ProviderName]\_[CollectionName]\_[YYYY-MM-DD].txt”. For example, `JSTOR_Arts&SciencesV_2008-12-01.txt`. Providers and recipients can agree in advance how best to present complex collection names.
- 5.3.1.5 All metadata should be provided as plain text. If metadata is provided in a format that does support additional style or formatting, it should be presented without those enhancements. Data should not include colors, typefaces, italics, or other markup.
- 5.3.1.6 Text should be encoded as UTF-8. The UTF-8 character set is well supported and encompasses the writing systems of many languages. This is also a common output option for programs such as Microsoft Excel.
- 5.3.1.7 One publication should be given in each line of the file, with a column for each field given in Section 5.3.2, Data Fields.
- 5.3.1.8 Data should be provided with column headers (see Section 5.3.2) and without a blank row between the column header and the first row of content.
- 5.3.1.9 A title should be listed twice if there is a coverage gap of greater than or equal to 12 months, with only the coverage field changing; greater granularity in reporting data coverage gaps is desirable, and should be agreed with the link resolver supplier if it can be supported.
- 5.3.1.10 All rows should be consistent in terms of format. For example, ISSN should always be expressed as nine characters with a hyphen separator, and date fields should always be in the format described in Section 5.3.2.
- 5.3.1.11 The metadata file should be supplied in alphabetical order by title to ensure ease of checking and import by knowledge base developers.

### 5.3.2 Data Fields

#### 5.3.2.1 Field and Labels

The content provider should include the following fields as columns in the tab-separated metadata file. *All fields should be considered to be mandatory if they exist*, and all effort should be made to gather the data, even if it must be obtained from another area of the business or even from an external source.

Because recipients of metadata files will be expecting to receive all files in a matching format, every field should appear in the order given below, even if the content provider is unable to provide any information, or no information is appropriate for a specific field. The following field labels should be used. To avoid

## KBART Phase I Recommended Practice

confusion and unnecessary errors, content providers are encouraged to include labels on every file they generate.

For consistency, the following labels should be used:

publication_title	Publication title
print_identifier	Print-format identifier (i.e., ISSN, ISBN, etc.)
online_identifier	Online-format identifier (i.e., eISSN, eISBN, etc.)
date_first_issue_online	Date of first issue available online
num_first_vol_online	Number of first volume available online
num_first_issue_online	Number of first issue available online
date_last_issue_online	Date of last issue available online (or blank, if coverage is to present)
num_last_vol_online	Number of last volume available online (or blank, if coverage is to present)
num_last_issue_online	Number of last issue available online (or blank, if coverage is to present)
title_url	Title-level URL
first_author	First author (for monographs)
title_id	Title ID
embargo_info	Embargo information
coverage_depth	Coverage depth (e.g., abstracts or full text)
coverage_notes	Coverage notes
publisher_name	Publisher name (if not given in the file's title)

Further details about the contents of each field are given below. Examples of complete records for various types of content are given Appendix A: Data Exchange Samples.

### 5.3.2.2 Publication Title

Give the full name of the publication, for example, as it appears on the print edition or on its web homepage. Special characters should be encoded using the UTF-8 character set. Abbreviations should be avoided.

Leading articles in a title should be included; for instance, "The Holocene" should be listed as "The Holocene" in its complete official form, not "Holocene."

Previous titles of the journal should be listed as separate entries, with their own set of coverage dates corresponding to the period of time in which that title was used. Knowledge base developers should then ensure appropriate matching between related titles.

Collection titles should not be given as individual titles within metadata files. Any collections of titles (packages) should be sent as a separate file with the collection name identified in the filename.

## KBART Phase I Recommended Practice

### 5.3.2.3 Print-format Identifier (e.g., ISSN, ISBN, etc.)

Provide the content's standard identifier. Initially, this will likely be the ISSN (presented with all 9 characters, including the hyphen and the check digit) or ISBN (either ISBN-10 or ISBN-13, as available; link resolvers can convert as necessary). In the future, this may include the ISMN, ISAN, and others.

In cases where multiple ISSN or ISBNs exist for the title, only the print-format ISSN or ISBN should be used in this field.

### 5.3.2.4 Online-format identifier (e.g., eISSN, eISBN, etc.)

In cases where identifiers for electronic formats have been created for the title, they should be included in this field.

### 5.3.2.5 Date of first issue available online

For journals, this field should include the date of the first issue available online, in the format YYYY-MM-DD. Use only those fields that apply; for example, if the journal is annual, only YYYY should be used, whereas if the journal is monthly or quarterly, only YYYY-MM should be used. Only in cases where issues of the journal have specified cover dates including the day should YYYY-MM-DD be used.

For books, the publication date should be given in the format YYYY-MM-DD. Again, use only those fields which are specifically given in the book's publication date.

The ISO 8601 date format should be used for all dates.

### 5.3.2.6 Number of first volume available online

For journals, give the volume number of the first issue in this field. Do not include any labels (e.g., "vol." or "v."). Try to reflect the house style for citing content, and give an alphanumeric value in this field if appropriate.

Knowledge base developers should use an equivalent logic to normalize this data and the data provided in OpenURL queries to maximize the likelihood of a citation being matched to a source.

For books, leave this field blank.

### 5.3.2.7 Number of first issue available online

For journals, give the issue number of the first issue. Do not include any labels (e.g., "no." or "n."). Do not include supplement or part values. You should try to reflect the house style for citing your content, and may give an alphanumeric value in this field if appropriate.

Knowledge base developers should use an equivalent logic to normalize this data and the data provided in OpenURL queries to maximize the likelihood of a citation being matched to a source.

For books, leave this field blank.

### 5.3.2.8 Date of last issue available online (or blank, if coverage is to present)

For journals, indicate the date of the most recent issue available. Again, use only those fields which are specifically given in the journal's cover date.

For journals, this field will be left blank if the journal is available "to the present."

## KBART Phase I Recommended Practice

For monographs, this field will always be blank.

### **5.3.2.9 Number of last volume available online (or blank, if coverage is to present)**

For journals, the volume number of the latest issue should be given in this field. Do not include any labels (e.g., “vol.” or “v.”). You should try to reflect the house style for citing your content, and may give an alphanumeric value in this field if appropriate.

Knowledge base developers should use an equivalent logic to normalize this data and the data provided in OpenURL queries to maximize the likelihood of a citation being matched to a source.

For journals, this field will be left blank if the journal is available “to the present.”

For books, leave this field blank.

### **5.3.2.10 Number of last issue available online (or blank, if coverage is to present)**

For journals, give the issue number of the latest issue. Do not include any labels (e.g., “no.” or “n.”). Do not include supplement or part values. You should try to reflect the house style for citing your content, and may give an alphanumeric value in this field if appropriate.

Knowledge base developers should use an equivalent logic to normalize this data and the data provided in OpenURL queries to maximize the likelihood of a citation being matched to a source.

For journals, this field will be left blank if the journal is available “to the present”.

For books, leave this field blank.

### **5.3.2.11 Title-level URL**

Indicate the URL of the title’s homepage. For journals, this page should be a listing of the available volumes and issues. For books, this page should be a table of contents.

### **5.3.2.12 First author (for monographs)**

For books, give the last name of the book’s first author.

For journals, leave this field blank.

### **5.3.2.13 Title ID**

Give the proprietary identifier for the content title, if you use a Title ID to create links to content. If more than one identifier exists, then supply the Title ID used for linking. If outside parties will not need to know or use your proprietary identifiers, or if no proprietary identifiers exist, this field may be left blank, but it would be preferable to include a titleID if one exists.

### **5.3.2.14 Embargo**

The embargo field reflects limitations on when resources become available online, generally as a result of contractual limitations established between the publisher and the content provider. Presenting this information to librarians (usually via link resolver owners) is vital to ensure that link resolvers do not generate links to content that is not yet available for users to access.



## KBART Phase I Recommended Practice

One of the biggest problems facing members of this supply chain is that multiple kinds of embargoes exist—in some cases, coverage “to one year ago” means that data from 365 days ago becomes available today, while in other cases it means that the item is not available until the end of the current calendar year.

Because of the complexities of embargoes, we recommend that the ISO 8601 date syntax should be used. This is flexible enough to allow multiple types of embargoes to be described.

The following method for specifying embargoes is derived from the ISO 8601 “duration syntax” standard, making a few additional distinctions not covered in the standard. The embargo statement has three parts: type, length, and units. These three parts are written in that order in a single string with no spaces.

**Type:** All embargoes involve a “moving wall,” a point in time expressed relative to the present (e.g., “12 months ago”). If access to the journal *begins* at the moving wall, the embargo type is “R”. If access **ends** at the moving wall, then the embargo type is “P”.

**Length:** An integer expressing the length of the embargo

**Units:** The units for the number in the “length” field: “D” for days, “M” for months, and “Y” for years. For simplicity, “365D” will always be equivalent to one year, and “30D” will always be equivalent to one month, even in leap years and months that do not have 30 days.

The “units” field also indicates the *granularity* of the embargo, that is, how frequently the moving wall “moves.” For example, a newspaper database may have a subscription model that gives customers access to exactly one year of past content. Each day, a new issue is added, and the issue that was published exactly one year ago that day is removed from the customer’s access. In this case, the embargo statement would be “R365D”, because the date of the earliest accessible issue changes each day.

Another journal may have a model that gives access to all issues in the current year, starting in January. The following January, the customer loses access to all of the previous year’s issues at once, and will only be able to access issues published in the current year. In this case, we would say that the customer has access to one “calendar year” of content. The embargo statement would be “R1Y”, because the date of the earliest issue changes once a year.

Below are some common embargoes expressed according to this syntax:

- Access to all content, except the current calendar year: P1Y
- Access to all content in the previous and current calendar years: R2Y
- Access to all content from exactly 6 months ago to the present: R180D
- Access to all content, except the past 6 calendar months: P6M

In the case where there is an embargo at both the beginning and end of a coverage range, then two embargo statements should be concatenated, the starting embargo coming first. The two statements should be separated by a semicolon. For example, “R10Y;P30D” describes an archive in which the past 10 calendar years of content are available, except for the most current 30 days.

## KBART Phase I Recommended Practice

### 5.3.2.15 Coverage depth

This field will indicate the extent to which content is covered within the range given in the coverage and embargo fields. It can have one of three values:

- **fulltext:** Indicates that the full text of articles is available. If the full text does not match the print equivalent, the “coverage notes” field can describe what is excluded (e.g., “excludes graphics”)
- **selected articles:** Coverage includes the full text of some, but not all articles. The specifics of the coverage policy may be indicated in the “coverage notes” field.
- **abstracts:** Coverage includes only abstracts of articles.

“Selected articles” should be used in this field only if a significant number of articles are omitted, perhaps as a result of specific policy. For example, a particular journal may only publish research articles online, and not letters or book reviews. Other databases may only include articles in certain subject areas. That policy should be described in the “coverage notes” field. The coverage depth should not be set to “selected articles” in cases where only a few articles are missing due to technical issues.

The above coverage depth values can be used in combination with a semicolon to delimit values. For example, if coverage of a journal includes only abstracts of selected articles (e.g., as may occur in A&I databases), this field would include “abstracts; selected articles”. A topic-oriented full-text product would be designated as: “selected articles”.

### 5.3.2.16 Coverage notes

This is an optional free-text field that may be supplied if the coverage depth used requires further explanation. This field is used to describe the specifics of the coverage policy, for example, “Excludes letters and book reviews.”

This field can be displayed verbatim in the link resolver results set so that end users can identify exclusions in content.

### 5.3.3 Error Reporting

When librarians and their users locate errors in knowledge base data (for example, incorrect coverage dates for a journal in a full text aggregation), we encourage them to report these errors to the knowledge base developer, who is expected to investigate the error and update the global knowledge base. The investigation time could take hours or days, depending on the level of cooperation of the original content provider. (Note that knowledge base developers do not have access to all electronic resources, and may be unable to confirm the error or determine the correct information if they do not have support from the content provider. In addition, depending on the link resolver vendor’s distribution model, it may take some time before the customer’s knowledge base reflects the change.)

The NISO/UKSG KBART Working Group recommends that knowledge base developers actively seek solutions that will allow corrections to the global knowledge base to appear locally in a much more timely manner, and thus eliminate the need for librarians to customize their collections to overcome shortcomings in the global knowledge base.

In a similar vein, we recommend the creation of a definition of what constitutes sufficient evidence to describe corrected metadata.

## 6 Education

---

Many of the problems with OpenURL technology stem from a lack of awareness of the technology, how it is used, and how it can benefit everyone in the supply chain. The NISO/UKSG KBART Working Group intends to help raise awareness and encourage participation by using various channels, including this Recommended Practice, a web-based information hub, PR, and training.

---

### 6.1 Web Hub

A central web hub—<http://www.uksg.org/kbart>, with information mirrored at <http://www.niso.org/workrooms/kbart>—will provide an authoritative starting point for those who need to learn about the OpenURL supply chain. The web hubs are optimized for search engines to ensure they are easily discoverable by those who most need it—and who may not be familiar with the core terms that would usually be used to access its content.

KBART's work is particularly focused on those organizations and individuals with a limited understanding of the OpenURL supply chain, so materials produced in support of its education mandate will assume existing knowledge and experience are minimal.

The hub breaks out the contents of this report into smaller, bite-sized modules, and will be supplemented by additional materials, listed below.

#### 6.1.1 Best Practice Guidelines

A summary of the metadata structure and exchange guidelines given in this Recommended Practice.

#### 6.1.2 FAQ

A set of frequently answered questions to help address common queries, as given in Section 6.3.

#### 6.1.3 Glossary

Definitions of core terms, as provided at the end of this Recommended Practice.

#### 6.1.4 Quick Guides / Fact Sheets

Information on:

- KBART summary: goals, benefits
- Supply chain summary: summarizing stakeholders and their responsibilities, exploring how information and technologies are used, explaining the benefits inherent in participation, and identifying the various products and tools available.

#### 6.1.5 Video demonstrations

Video demonstrations showcasing OpenURL technology and knowledge base metadata in practice.

#### 6.1.6 Case studies

A set of case studies demonstrating good practice in transfer of metadata to knowledge bases.

### 6.2 Outreach

To raise awareness of KBART's outcomes and output, PR activities have been, and will continue to be, undertaken. Those include press announcements, publishing articles, speaking engagements, and participating in training seminars/webinars. Where possible, we will leverage the existing reach of organizations such as ALPSP, ER&L, LITA, ACRL, ALA, NASIG, NFAIS, NISO, OASPA, PSP, PA, STM, UKSG, VALA, and others.

---

### 6.3 FAQ Examples

A frequently asked questions (FAQ) section of the web hub will be useful for those starting to work through the OpenURL, or those seeking more specific information. Examples of questions that will appear in the FAQ are listed below.

#### 6.3.1 General Questions

- What is a link resolver?
- What is a knowledge base?
- What is a link source/link target... and what's the difference?
- What is an OpenURL?

#### 6.3.2 Content Provider/Publisher Questions

- How do I make sure my customers can link to my content?
- How do I help my customers link from my site to other content?
- Who maintains knowledge bases and how do I contact them?
- How is the information about my content displayed to the user of a library system?
- What kind of information does a knowledge base need in order to ensure my customers see my content?
- Do all knowledge base developers need the information in the same way?
- Should I include information about my package offerings?
- If I supply knowledge base developers with information, do I still need to give holdings and linking information directly to librarians?
- I have provided information on my content, but I am told customers still can't reliably link to my content. Why is this?
- Can I provide data about my e-books as well as my journals?
- How often should I provide new information about my content?
- When describing the range of content available, should this be done with dates or volume/issue ranges?
- Do I need to provide information on all the content available, or only on the content that is licensed? What if there are different license models?
- My customers are registering their link resolver prefixes and they all look slightly different. Is there any standard I should expect for the entry of this information and how can I validate whether they are correct?

## KBART Phase I Recommended Practice

### 6.3.3 Librarian Questions

- What are the benefits of having a link resolver?
- What information do I need to supply to my link resolver?
- Why can't content providers provide reliable information on their holdings? I licensed what I was told was a standard collection from Publisher X, but I find that I still need to customize the holdings information.
- Is there a list of OpenURL-compliant (as in, I can register my link resolver) platforms available?
- Which is the best source of accurate data about my holdings?
- Can I upload data to the knowledge base, or must changes be made manually?
- Do all link resolvers use the same collection profiles and knowledge base? If not, how can I assess the coverage of different services?
- When a journal changes title, do I need to amend my holdings in the knowledge base, or is this done automatically?
- When a journal changes publishers, do I need to amend my holdings manually?
- When I make changes to my holdings, does the knowledge base update dynamically?
- I have just licensed a new full-text database or a new e-journal, but it does not appear in the knowledge base. Why not, and what should I do?
- Can I include information about print journal holdings in my knowledge base?
- Is it possible to customize the interface of my knowledge base? For example, can we provide links to other library services, like interlibrary loan?

---

### 6.4 Guidelines on Improved Linking

One problem facing some linking usage is badly formatted and poorly implemented link resolving. OpenURLs that are improperly formatted will invariably lead to a failure to successfully resolve to an appropriate location. Link resolvers generally transfer the information they receive with minimal checking of the metadata they are given from the source. For instance, the link resolver may check to ensure that the ISSN is properly formed (though not necessarily check to ensure that it is the *correct* ISSN for the citation given), and then delete it if the check digit fails. Another example might be that the resolver may take the provided title and then return a more accurate or more standardized version of the title.

But if the source link provider has created an OpenURL in which the enumeration data (e.g., "volume," "issue," "number," etc.) is in the place of the chronology data (e.g., day, month, year, etc.), and vice versa, the resolver will not reverse and correct that information. Therefore, it is vital that the organization creating the source OpenURL link do so accurately and correctly. A number of resources exist online to help such organizations create these links. Examples include:

## KBART Phase I Recommended Practice

- COinS Generator  
<http://generator.ocoinc.info/>
- OpenURL Generator Help  
<http://search.lib.unimelb.edu.au/help/openurlgen.html>

In addition, a project funded by The Andrew W. Mellon Foundation and based at Cornell University is exploring the possibility of a system that will analyze and report on a large number of source OpenURLs in order to identify structural errors and assist source OpenURL generators in correcting these problems. Adam Chandler and David Ruddy of the Cornell University Library are currently engaged in a Mellon Planning Grant project with Eric Rebillard, Professor of Classics at Cornell. Professor Rebillard is editor of *L'Année philologique*, a citation database. Since 2004, the electronic version of *L'Année philologique* (called APh Online) has been OpenURL compliant. Each record contains a link that can be processed by a link resolver, but many OpenURL links fail. Adam Chandler developed initial recommendations for metadata improvement based on a manual review of a sample set of 126 OpenURLs generated by *L'Annee*. His report, "Results of *L'Année philologique online* OpenURL Quality Investigation," identified many typical metadata problems that cause OpenURLs to fail: malformed dates, volume and issue numbers combined into one field, start page fields with entire page ranges, lack of identifiers, etc. Such a report is extremely useful to *L'Annee* because it precisely identifies the critical failure points where improvement efforts can most profitably be focused. However, performing such a manual review of all the OpenURLs generated by *L'Annee*, or any other vendor, would be prohibitively expensive and time consuming.

Chandler and Ruddy proposed exploring the feasibility of developing a fully automated OpenURL evaluation process. Such a system would accept OpenURLs and return scores based on a set of evaluation metrics. These scores would allow resource providers to see precisely where their OpenURLs were weakest, letting them target metadata improvement efforts in the most cost-effective manner. They ultimately envision a community recognized index for measuring the quality of OpenURL links from content providers. In December 2009, NISO approved a two-year project to focus on the testing and validating of the metrics used to determine OpenURL quality, to be chaired by Chandler (see <http://www.niso.org/workrooms/openurlquality> for more information).

## 7 Next Steps /Phase II

---

While this Recommended Practice outlines a number of important steps that content providers can take to improve how data is transferred among and between the members of the e-resources supply chain, the NISO/UKSG KBART Working Group believes there are many additional steps that can be taken by all stakeholders to further improve the library user's experience when using link resolvers and their related knowledge bases. These are as follows:

- Definitions for global vs. local updates (see Section 7.1.1)
- Consortia-specific metadata transfer
- Institution-specific metadata transfer
- Documentation of guidelines for non-text content metadata transfer
- Review of metadata transfer for e-books

## KBART Phase I Recommended Practice

- Monitoring and enforcing compliance with KBART recommendations
- Exchange of ERM data

The KBART Working Group welcomes feedback on the future direction of KBART from all users of this Recommended Practice.

---

### 7.1 Recommendations and Further Discussion

#### 7.1.1 Differences Between Aggregations and Individual Content Providers

It is important to note that coverage issues for e-journals and e-journal packages are very different from coverage issues for a title within an aggregated database. All subscribers of an aggregation have the same rights to the same titles; libraries are simply leasing access to the collections created by database aggregators. When content is added or deleted, it is added or deleted for *all* institutions that subscribe to that package, and is very often added or deleted without the institution's knowledge. As a result, when a coverage error is found and the correct data is verified, the error can be fixed—once—in the global knowledge base.

For individually subscribed e-journals, however, including those in e-journal packages, the title lists and the access rights for the titles often vary from institution to institution, and from contract to contract. The global knowledge base can only represent the overall dates of content available for the title on the online host. The access rights or coverage entitlements must be managed on a customer-by-customer basis, usually within either a customized version of the master knowledge base or a locally installed copy of the knowledge base that is customized by the institution. Coverage discrepancies discovered for an e-journal by one customer may not apply to other customers; as a result, the institution must update just its own holdings list, and not the master knowledge base.

An overview of the current supply chain for link resolver knowledge base data should indicate some of the many challenges facing those trying to make the most of electronic resources via the OpenURL.

#### 7.1.2 Consortial Package Challenges

Institutions that purchase some or all of their electronic resources through library consortia have additional challenges in the transfer of metadata due to complexities introduced by the consortial purchasing. Library consortia differ greatly: some may simply offer a pricing discount for an existing and static product, while others may create unique collections of e-journal content for members of the consortium. Still others may provide a “top-up” purchasing opportunity, whereby members can gain access to resources they had not previously purchased.

When the consortium influences selection of e-resources, it is beneficial for the consortium to be able to distribute its purchasing specifics via the knowledge base manager to all those who have access to these resources.

#### 7.1.3 Institution-specific Metadata Transfer

In addition to the challenges faced in describing consortial packages, individual institutions often need to localize the global target in order to make coverage statements and active titles relevant to their library holdings. KBART intends to address procedures for metadata transfer of localized holdings data in a future phase.

## **KBART Phase I Recommended Practice**

### **7.1.4 Non-text Content Metadata Transfer**

So far, KBART has concentrated on guidelines for metadata transfer of text content (e-journals and e-books). However, populating knowledge bases with metadata describing non-text content is becoming increasingly important.

### **7.1.5 Review of Metadata Transfer for E-Books**

The procedures documented in this report allow for the creation and transfer of files describing both e-journal and e-book holdings. The existing guidelines need to be reviewed in light of any feedback from stakeholders on improvements to data fields for e-book content.

### **7.1.6 Exchange of ERM Data**

So far, KBART has addressed the problem of populating link resolver knowledge bases with holdings data. In a future phase, the group will consider the additional data required to populate ERM knowledge bases with data relating to e-resource subscriptions.

### **7.1.7 Compliance with KBART Recommendations**

At the moment, the KBART Working Group members do not feel that we should be working toward a structured standard regarding this data. As noted in Section 5, there are acceptable alternatives to many of the recommendations made in this report, and we do not wish to create extra work for organizations that already have an effective method for distributing metadata. However, we do wish to raise awareness of what practices can be beneficial or problematic when distributing metadata, and hope that our recommendations can be used by organizations to improve communication with other parts of the supply chain.

To this end, the NISO/UKSG KBART Working Group recommends that this KBART Recommended Practice form a code of practice—similar to that of Project TRANSFER—that organizations are encouraged to endorse. We will also consider establishing a public list of content providers and knowledge base developers that can send and use data in the recommended format.



## KBART Phase I Recommended Practice

### Appendix A: Data Exchange Samples

---

The following example displays a complete holdings file for a single package from The Royal Society. In this example, the content provider only offers serials, so the identifiers shown in the print\_identifier field are only ISSN, not ISBNs. The data itself should be exchanged in a tab-delimited format, but the data is presented here in a table, solely to make the data layout more understandable for readers. Our thanks to The Royal Society for allowing us to use this file as an example.

publication_title	print_identifier	online_identifier	date_first_issue_online	num_first_vol_online	num_first_issue_online	date_last_issue_online	num_last_vol_online	num_last_issue_online	title_url	first_author	title_id	embar_go_info	coverage_type	covergae_notes	publi_sher_name
Philosophical Transactions	0370-2316		1665	1		1678	12		<a href="http://rstl.royalsocietypublishing.org/">http://rstl.royalsocietypublishing.org/</a>		rstl		fulltext		
Philosophical Transactions	0260-7085		1683	13		1775	65		<a href="http://rstl.royalsocietypublishing.org/">http://rstl.royalsocietypublishing.org/</a>		rstl		fulltext		
Philosophical Transactions of the Royal Society of London	0261-0523		1776	66		1886	177		<a href="http://rstl.royalsocietypublishing.org/">http://rstl.royalsocietypublishing.org/</a>		rstl		fulltext		
Philosophical Transactions of the Royal Society of London. (A.)	0264-3820		1887	178		1895	186		<a href="http://rsta.royalsocietypublishing.org">http://rsta.royalsocietypublishing.org</a>		rsta		fulltext		
Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character	0264-3952		1896	187		1934-01-01	233		<a href="http://rsta.royalsocietypublishing.org">http://rsta.royalsocietypublishing.org</a>		rsta		fulltext		

### KBART Phase I Recommended Practice

Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences	0080-4614		1934-11-09	234	731	1990-06-30	331	1622	<a href="http://rsta.royalsocietypublishing.org">http://rsta.royalsocietypublishing.org</a>		rsta		fulltext		
Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences	0962-8428		1990-07-16	332	1623	1995-12-15	353	1703	<a href="http://rsta.royalsocietypublishing.org">http://rsta.royalsocietypublishing.org</a>		rsta		fulltext		
Abstracts of the Papers Printed in the Philosophical Transactions of the Royal Society of London	0365-5695		1800	1		1843	4		<a href="http://rspl.royalsocietypublishing.org/">http://rspl.royalsocietypublishing.org/</a>		rspl		fulltext		
Abstracts of the Papers Communicated to the Royal Society of London	0365-0855		1843	5		1854	6		<a href="http://rspl.royalsocietypublishing.org/">http://rspl.royalsocietypublishing.org/</a>		rspl		fulltext		
Proceedings of the Royal Society of London	0370-1662		1854	7		1905-01-01	75								
Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character	0950-1207		1905-04-22	76	507	1934-10-15	146	859	<a href="http://rspa.royalsocietypublishing.org">http://rspa.royalsocietypublishing.org</a>		rspa		fulltext		

### KBART Phase I Recommended Practice

Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences	0080-4630		1934-11-01	147	860	1938-02-18	164	919	<a href="http://rspa.royalsocietypublishing.org">http://rspa.royalsocietypublishing.org</a>		rspa		fulltext		
Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences	0080-4630		1938-03-18	165	920	1969-01-28	308	1495	<a href="http://rspa.royalsocietypublishing.org">http://rspa.royalsocietypublishing.org</a>		rspa		fulltext		
Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences	0080-4630		1969-02-18	309	1496	1990-06-08	429	1877	<a href="http://rspa.royalsocietypublishing.org">http://rspa.royalsocietypublishing.org</a>		rspa		fulltext		
Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences	0962-8444		1990-07-09	430	1878	1995-12-08	451	1943	<a href="http://rspa.royalsocietypublishing.org">http://rspa.royalsocietypublishing.org</a>		rspa		fulltext		

## Glossary of Terms

---

### **Aggregator**

A bibliographic service that provides online access to the digital full text of periodicals published by different publishers. Subscriptions are available by package, rather than title by title. Typically packages vary by the type of library (e.g., special, academic, public). See also: **full-text host**.

### **Appropriate copy**

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

### **Article-level link**

A URL that takes a user directly to the correct article they seek, rather than to the publication, volume, or issue in which the article is published. An article-level link may take the user to an abstract or to a version of the full text if it is available to the user. The article level link may be embodied in an OpenURL, but is not necessarily so.

### **Consortia**

Collections of libraries that work together to purchase and provide access to resources; in some cases, these are simply “buying clubs,” whereas in others they are a closely integrated network of related libraries. Publishers also join together into consortia to offer cross-publisher content packages to libraries.

### **Content packages**

Bundles of content that can be purchased more cheaply than content purchased via separate licenses to the individual components; for example, packages offered by single-publisher “big deals,” aggregators, or publisher consortia.

### **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.

### **Context sensitive**

Context sensitive means that the destination of a link is determined by the user’s context, i.e., the institution with which the user is affiliated, and the preferences registered by that institution.

### **Customization**

The process by which an institution customizes information contained in the link resolver software’s global knowledge base in order to reflect its local collections and services. This may include restrictions on holdings, type of service, and departmental or group access designations. Also commonly referred to as “localization.”

### **Document delivery**

Provision of individual documents upon request and payment (either immediate or through a pre-paid deposit account) from the reader or their representative, for use only by that reader. A document delivery service specializes in the transactions and is

## **KBART Phase I Recommended Practice**

generally not the publisher or primary full-text host of the content.

### **Digital Object Identifier (DOI)**

The Digital Object Identifier (DOI) system uniquely defines a digital resource, be it a journal article, a book chapter, a paragraph, an image, or some other item. Content providers must be affiliated with a DOI registration agency, and must own the rights to the content, in order to assign DOIs. See also ANSI/NISO Z39.84-2005, *Syntax for the Digital Object Identifier*.

### **DOI resolution**

DOIs are held in a database alongside the metadata describing the object they identify, and a current URL for that object. DOIs can be resolved by appending them to the base URL <http://dx.doi.org>. This directs them to the central DOI database, which looks up the current URL associated with the object, and redirects the user's browser to that location. CrossRef (<http://www.crossref.org/>) is the official DOI link registration agency for scholarly and professional publications.

### **Embargo**

A limitation on access to a resource, placed by the publisher on distributors of the publisher's data, usually to prevent the cancellation of individual subscriptions. For example, a publisher's own website provides current issues of their e-publications, but an aggregator's website only provides issues older than one year.

### **Enumeration**

The use of number, volume, and issue descriptors to identify a specific journal issue, as opposed to the use of chronological descriptors.

### **Electronic Resource Management (ERM)**

A broad term for a collection of tools to help libraries manage their electronic resources.

### **Free content**

Content that can be accessed by any individual with internet access. The individual may need to register in some way to access the content, but they do not need to pay for the content or access to the content.

### **Full-text host**

A vendor who is contracted by the publisher to host the full text of publications in a single, searchable database to which access is enabled by subscriptions to individual titles or via article document delivery, rather than via a license to the entire database or parts thereof.

### **Gateway**

A site that channels users to full text without hosting that full text; for example, a subscription agent's website that authenticates the user and directs them to the full text at a publisher's website.

### **Integrated Library System (ILS)**

A collection of integrated tools for managing all the different parts of a library's collection management system.

### **Inbound linking (syntax)**

Links into a website from other online resources. A content provider is enabling inbound linking if it makes 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).

## KBART Phase I Recommended Practice

### **Indexes, abstracts, and full-text content**

Indexes contain article title, author, and bibliographic information (e.g., journal title, volume, issue, year, and page numbers) for journal content. An “abstracting and indexing” (A&I) database contains this information along with each article’s abstract. Full-text content is the entire text of the article, and is typically only available from the publisher or other licensed content providers/aggregators.

### **Knowledge base**

An extensive database maintained by a knowledge base developer that contains information about electronic resources, such as title lists, coverage dates, inbound linking syntax, etc. The knowledge base 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.

### **Knowledge base developer**

An organization that compiles, distributes, and maintains the knowledge base.

### **License**

A contractual agreement between a content provider and a library or consortium. Among other specific clauses, the license determines the range of content (e.g., user can access volume X to volume Y) available for online access, and may also define a period of access (e.g., the license is active for three years, after which time all access is terminated). These terms may vary for each organization and should be reflected in the local knowledge base.

### **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 (see ANSI/NISO Z39.88-2004, *The OpenURL Framework for Context-Sensitive Services*) to enable inbound linking; others base their link-to syntax on proprietary, but predictable, identifiers.

### **Localization**

The process by which an institution customizes information contained in the link resolver software’s global knowledge base in order to reflect its local collections and services. This may include restrictions on holdings, type of service, and departmental or group access designations.

### **Metadata**

Data about data—that is, information that describes content. For an article, this might be its title, the names of its authors, the title of the journal from which it is taken, and the volume, issue, and article or page numbers.

### **Online Public Access Catalogue (OPAC)**

The public interface for a library’s catalogue; just one part of the ILS.

### **Open access**

Business model by which full-text content is free at the point of access, i.e., users do not need to pay for a subscription or other license to view full text.

## KBART Phase I Recommended Practice

### **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).

### **Pay-per-view**

Online payment in exchange for permission to read an individual document. This is the common means by which readers can obtain an individual article or book (or chapter of a book) if they or their organization do not have a subscription to the resource containing the document. This service is provided by publishers and full-text hosts.

### **SFX**

One of a number of commercially available link resolvers; SFX was developed by Herbert van de Sompel and then commercialized by Ex Libris. It was the first resolver available on the market. A list of other link resolvers available today can be found at: <http://www.loc.gov/catdir/lcpaig/openurl.html>.

### **Source**

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

### **Subscribed content**

Content that the current user is licensed to access.

### **Subscription agent**

A company that is contracted by publishers to sell subscriptions (or other types of access license) to libraries, consortia, and other institutions.

### **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.

### **Title-level links**

A URL that takes a user directly to the publication they seek, rather than to a specific volume, issue, or article within it. Title-level links are not necessarily in the format of an OpenURL.

## Bibliography

---

- Chandler, Adam. "Results of *L'Année philologique online* OpenURL Quality Investigation." In *Mellon Planning Grant Final Report*. February 2009. Available at: <http://metadata.library.cornell.edu/oq/files/200902%20lannée-mellonreport-openurlquality-final.pdf>
- CrossRef. Available at: <http://www.crossref.org>
- COinS Generator. Available at: <http://generator.ocoinc.info>
- Culling, J. *Link Resolvers and the Serials Supply Chain*. Oxford: Scholarly Information Strategies, 2007. Available at: <http://www.uksg.org/projects/linkfinal>
- ISSN International Centre. Available at: <http://www.issn.org>
- ISO 8601:2004 *Data elements and interchange formats -- Information interchange -- Representation of dates and times*. Geneva: International Organization for Standardization, 2004.
- Library of Congress. *US ISSN Center Homepage*. Available at: <http://www.loc.gov/issn>
- Library of Congress, Portal Applications Issues Group. *OpenURL Resolver Products & Vendors*. Available at: <http://www.loc.gov/catdir/lcpaig/openurl.html>
- National Information Standards Organization (U.S.). *ANSI/NISO Z39.84-2005, Syntax for the Digital Object Identifier*. Bethesda, Md.: NISO, 2005. Available at: <http://www.niso.org/standards/z39-84-2005/>
- National Information Standards Organization (U.S.). *ANSI/NISO Z39.88-2004, The OpenURL Framework for Context-Sensitive Services*. Bethesda, Md.: NISO, 2004. Available at: <http://www.niso.org/standards/z39-88-2004>
- NISO. *KBART (Knowledge Base And Related Tools) Working Group*. Available at: <http://www.niso.org/workrooms/kbart>. See also: <http://www.uksg.org/kbart>
- NISO. *OpenURL Quality Metrics Working Group*. Available at <http://www.niso.org/workrooms/openurlquality>
- OCLC (OpenURL Maintenance Agency). *Registry for the OpenURL Framework - ANSI/NISO Z39.88-2004*. Available at: <http://openurl.info/registry>
- The British Library. *ISSN UK Centre*. Available at: <http://www.bl.uk/bibliographic/issn.html>
- The University of Melbourne. *OpenURL Generator Help*. Available at: <http://search.lib.unimelb.edu.au/help/openurlgen.html>
- UKSG. *KBART: Knowledge Bases And Related Tools Working Group*. Available at: <http://www.uksg.org/kbart>. See also: <http://www.niso.org/workrooms/kbart>