

Knowledge Bases and Related Tools (KBART) Recommended Practice



*A Recommended Practice of the
National Information Standards Organization*

Prepared by the
KBART Phase II Working Group

About NISO Recommended Practices

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

This recommended practice may be revised or withdrawn at any time.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

Published by

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

Copyright © 2014 by the National Information Standards Organization

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

NISO, 3600 Clipper Mill Road, Suite 302, Baltimore, MD 21211.

ISBN: 978-1-937522-41-4

10 9 8 7 6 5 4 3 2 1

Contents

Foreword	v
Section 1: KBART after Phase I	1
1.1 Establishment of an Endorsement Process and KBART Registry	1
1.2 KBART Survey of Libraries and Consortia	1
1.3 KBART Outreach.....	2
1.4 KBART Information Hub	2
Section 2: Compliance and Collaboration	3
2.1 Compliance	3
2.2 Collaboration Activities	4
Section 3: Metadata for Consortia	6
3.1 Introduction	6
3.2 Consortia Survey Summary.....	6
3.3 Recommendations for Consortium-Specific Files.....	7
3.4 File Contents and Naming Changes	7
Section 4: Open Access Metadata	9
4.1 It's Free Stuff Isn't It? Why Is Metadata for OA Required?	9
4.2 Open Access Survey Summary	9
4.3 Consideration of Types of Open Access	10
4.4 Changes in the Recommendations.....	10
Section 5: E-Books and Conference Proceedings	12
5.1 Importance of Metadata.....	12
5.2 Changes of Title	13
5.3 Changes in the Recommendations.....	13
Section 6: Guidelines for Effective Exchange of Metadata with Knowledge Bases	16
6.1 Method of Exchange	16
6.2 Frequency of Exchange	16
6.3 Data Contacts	16
6.4 Data Format.....	16
6.5 File Name	17
6.6 KBART Data Fields	18
Section 7: Next Steps	27

Appendix A: Sample Files.....28

Appendix B: Survey Results 36

Bibliography.....71

Foreword

About This Recommended Practice

In 2007, UKSG, a nonprofit organization that connects the information community, commissioned a report, *Link Resolvers and the Serials Supply Chain*. 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 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 and propose a community solution, backed up by educational activities and an Information Hub to support adopters. The joint NISO/UKSG KBART (Knowledge Bases And Related Tools) Working Group was established in December 2007 and the Phase I Recommended Practice published in January 2010 (NISO RP-7-2010) was the result of the initial phase.

Original Charge and Scope

The KBART Working Group’s original charge was to improve the supply of data to link resolvers and knowledge bases, in order to improve the efficiency and effectiveness of OpenURL linking. This was 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.

The NISO/UKSG KBART Working Group’s scope focused on problems in the information supply chain that relate to the data supplied to knowledge bases. This specifically excluded 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 (R2010), *The OpenURL Framework for Context-Sensitive Services*), or within other NISO working groups. The group 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.

KBART Phase II

With the completion of Phase I, it became clear that there was a need to continue work on more specific improvements and the group membership was re-aligned to support this work. The KBART Phase II Working Group met between February 2010 and March 2013 and achieved the following:

- 1) Development of a revised recommended practice to build on the initial recommendations delivered for Phase I of the KBART project in order to effect smoother interaction between members of the knowledge base supply chain. Whereas the Phase I report provided minimum recommendations to reach this goal, the Phase II revision focuses on the more granular, complex issues that cause problems in metadata supply. Knowledge base providers and their customers (primarily academic libraries) will benefit from provision of higher-quality data by content providers. Publishers will benefit from accurate linking to their content and subsequently increased usage.
- 2) Provision of educational and outreach opportunities that will address the needs of all stakeholders in the supply chain, description of the functions each needs to carry out to improve supply of data to knowledge bases, and setting out the value of doing so in each case. Content providers will benefit from a greater understanding of the needs and activities of those to whom they supply data. Knowledge base providers and libraries will again benefit from improvements that can be

expected when content providers are better supported through best practice recommendations on metadata provision.

- 3) Delivery of a centralized information portal, to support educational activities and provide comprehensive resource for further information, including, but not limited to:
 - a) Glossary (already available on the UKSG and NISO websites from Phase I)
 - b) Entry-level explanation of OpenURL: purpose, methodology, value (already available in Phase I Recommended Practice)
 - c) Introduction to knowledge base supply chain parties: roles, needs, responsibilities, value (already available in Phase I Recommended Practice)
 - d) Endorsement framework for content providers and knowledge base providers
 - e) Guidelines for best practice: knowledge base format and terms of provision
 - f) Crossover work with other standards/initiatives/technologies, e.g., ONIX Serials Online Holdings (SOH)
 - g) Contacts Registry for the purpose of linking to metadata webpages and provision of technical contact details within content provider and knowledge base organizations

NISO D2D Topic Committee Members

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

Kristin Antelman
North Carolina State University Libraries

Christine Stohn
Ex Libris, Inc.

Pascal Calarco, Co-chair
University of Waterloo Library

Lucy Harrison, Co-chair
Florida Virtual Campus

Juli Marsh
The Library Corporation (TLC)

Chris Shillum
Reed Elsevier

Tim Shearer
University of North Carolina Chapel Hill
Libraries

KBART Phase II Working Group Members

The following individuals served on the KBART Phase II Working Group, which developed and approved this recommended practice:

Magaly Bascones, Co-chair
JISC Collections

Andreas Biedenbach, former Co-chair
Independent Professional

Chad Hutchens, Co-chair
University of Wyoming Libraries

Benjamin Johnson
Serials Solutions

Noah Levin
Springer

Sheri Meares
EBSCO Information Services

Paul Moss
OCLC Online Computer Library Center

Gary Pollack
Cengage Learning

Jason Price
Claremont Colleges Library

Sarah Price, former Co-chair
University of Birmingham

Rose Robinson
Publishing Technology plc

Liz Stevenson
University of Edinburgh

Christine Stohn
Ex Libris, Inc.

Ruth Wells
Publishing Technology plc

Laird Barrett
Taylor and Francis

Julie Zhu
American Institute of Physics (AIP)

The following individuals served on the KBART Working Group during part of the recommended practice's development, but were no longer with the group at time of final approval:

Louise Cole
Kingston University

Sherrard Ewing
Serials Solutions

Marieke Heins
Swets Information Services

Matthew Llewellyn
The Royal Society

Catherine Sharp
University College London

Maria Stanton
Serials Solutions

Sarah Wright
Taylor and Francis

Trademarks and Service Marks

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

Section 1: KBART after Phase I

KBART Phase I was completed with the publication of the KBART Phase I Recommended Practice in January 2010 (NISO RP-9-2010). The NISO/UKSG KBART Working Group then identified additional topics for discussion to further improve the library user's experience when using link resolvers and their related knowledge bases. NISO formed a new working group for KBART Phase II at the beginning of 2010 to continue this work. The key topics considered in Phase II were consortia-specific metadata transfer (see [Section 3](#)), review of metadata transfer for open access publications (see [Section 4](#)), review of metadata transfer for e-books and conference proceedings (see [Section 5](#)), and the setup of an endorsement framework to enable content providers to achieve compliance (see [Section 2](#)). Another important action activity was to raise awareness of KBART recommendations from Phase I by undertaking speaking engagements and writing article for various publications in the library and publishing community.

1.1 Establishment of an Endorsement Process and KBART Registry

A substantial contributor to full text usage on publisher sites is from library-mediated technology. A significant part of this library access flows through OpenURL link resolvers that rely on detailed knowledge base holdings data provided by the content publishers. It is therefore vital that this knowledge base information is accurate and up-to-date. The KBART Recommended Practice optimizes this process through provision of a file format and common sense set of metadata fields and formats for transmission of holdings metadata from content providers to link resolver knowledge bases. The standardized KBART file makes the file transfer and knowledge base updates easier to handle. In addition, some consortia are now expecting KBART-compliant files for their knowledge bases and are incorporating this requirement into their contracts with content providers. A key outcome of the early stages of the Phase II group activity was an endorsement framework. This allows content providers to achieve KBART compliance and publicly celebrate their commitment to good quality metadata provision.

All content providers, from major databases to small publishers, were encouraged to publicly endorse the Phase I KBART Recommended Practice by submitting a sample file to the KBART working group, at kbart@niso.org. Once the file's format and content was reviewed and approved, and the provider had made it publicly available (in line with the recommendations), the provider was added to a public list of endorsing providers. (With the publication of this Phase II Recommended Practice, an updated endorsement process will be put in place. More details about the new endorsement and compliance process can be found in section [2.1](#).)

In addition to the endorsement process, a contacts registry was made available for content providers and knowledge base developers to register their organization's advice for downloading holdings metadata as well as providing their metadata webpage and contact details. All endorsing providers that have been approved for KBART compliance carry the logo next to their name in the registry. The KBART Working Group is proud of the take-up within the last two years and is constantly encouraging additional content providers to endorse KBART and submit their files for verification.

1.2 KBART Survey of Libraries and Consortia

Feedback from the community is very important for all of KBART's work but is even more crucial for the specific problem areas that Phase II is concerned with. The KBART group conducted a SurveyMonkey survey in January 2012 to obtain information about the use of consortia title lists and about the library and publishing community's views on the metadata of Open Access material. The survey—announced on various listservs, including LIS-E-Resources, SERIAL-ST, ICOLC, Web4Lib, and LIS-NESLI-Reps—

generated over 200 responses. The summary results of the survey can be found in [Section 3, Metadata for Consortia](#), and [Section 4, Open Access Metadata](#). The detailed survey questions and responses are in [Appendix B](#).

1.3 KBART Outreach

On publication of the KBART Phase I Recommended Practice in January 2010, the project group increased its profile in speaking at relevant events in order to inform the library and publishing community on the output of the working group.

As a result, KBART representatives were invited to speak at various conferences and meetings, and to provide articles for publications. One of the highlights was an extended report about KBART in *Against the Grain* in February 2011. A number of additional articles were also published and numerous speaking engagements were made. A list of these with links to available full text or presentation slides is available from the KBART website (www.niso.org/workrooms/kbart/resources/).

The group will present the Phase II recommendations at various conferences beginning in late 2013.

The KBART group has also set up a mailing list for all interested parties within the library community. Subscribe by sending an e-mail to kbart_interest-subscribe@list.niso.org. Subscribers will receive periodic announcements and e-mails about group activities and can initiate discussions with others on the list to ask questions or share experiences in using the KBART recommended practices. At the time of writing, the list has more than 300 subscribers.

1.4 KBART Information Hub

The KBART Information Hub on the NISO website (www.niso.org/workrooms/kbart) provides additional information related to the KBART recommended practices including:

- KBART Glossary
- Endorsement information
- Registry of knowledge base supply chain contacts
- Background information on OpenURL and knowledge bases

Section 2: Compliance and Collaboration

2.1 Compliance

The KBART support process includes a framework for enabling content providers to achieve compliance and publicly endorse their commitment to providing high quality metadata to the supply chain, which is described in detail below.

2.1.1 Phase II Endorsement

The publication of the Phase II Recommended Practice supersedes the Phase I Recommended Practice. This means that from date of publication of this Phase II RP, all new content providers that contact the KBART group will be working towards Phase II compliance and will be endorsed as such. Content providers who were endorsed under Phase I should consult section [2.1.4](#) below.

It is acknowledged that not every field in the Phase II RP will be relevant to all content providers. For example, fields specific to monographs and conference proceedings will not be applicable to purely journal metadata. However, 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 an external source.

2.1.2 Requesting Endorsement

Knowledge base vendors can endorse the KBART Phase II Recommended Practice by confirming that their systems can process KBART formatted files. The steps for requesting endorsement are listed in this section.

- 1) Read the requirements, accessible via www.niso.org/workrooms/kbart/endorsement/, which also includes a link to a sample data file.
- 2) Format your e-journal and e-book holdings metadata to meet the requirements.
- 3) Indicate your interest in endorsement by contacting the KBART working group at: kbart@niso.org and providing the following sample files:
 - A sample for the complete A-Z global list
 - A sample for a package within the global A-Z list (e.g., STM collection) that is based on how the package is sold
 - A sample for a consortium-specific package (if collections have specific differences within regional consortia)
- 4) Ensure that you have a process in place for regular data exchange as outlined in section [6.1](#).

2.1.3 Verification for Endorsement

Knowledge base vendors in the KBART Working Group will test the sample file(s) provided by content providers who are requesting endorsement. This testing process will verify that:

- 1) The filename is correct and is relevant to the level (e.g., package / consortium)
- 2) The KBART structure within the file is adhered to and data is in the correct position
- 3) The structure guidelines within each field are adhered to (e.g., correct use of date format)
- 4) All data in mandatory fields are included (e.g., identifiers, title, URLs)

The KBART Working Group will also sample check for known issues such as how former titles are represented.

In addition, the content provider is also required to perform the following activities to achieve KBART compliance:

- 1) Post files on a publicly available website with the ability to download and re-use.
- 2) Ensure that the file represented is a comprehensive picture of coverage for the global A-Z list / package / region / consortium as appropriate.
- 3) Keep the files up-to-date with all holdings changes, e.g., title transfers, cessations, and coverage changes.
- 4) Overwrite older files with a completely new file that will allow the knowledge base vendor to download and compare with the current knowledge base file.

Once the content provider's files are approved by the KBART Working Group testers:

- 1) The content provider should register its organization on the KBART contacts registry website (<https://sites.google.com/site/kbartregistry/>), providing a link to a publicly available webpage where it will be possible to download the newly KBART formatted dataset(s). The metadata files should be made available in human-readable format.
- 2) The KBART Working Group will provide a logo for the content provider webpage.
- 3) The content provider can distribute its KBART formatted data (or its download location) to their supply chain contacts.

With the above compliance verification, stakeholders can be assured that metadata is trusted and activated within a global knowledge base at the appropriate level of granularity without the burdensome task of title-by-title checking. This ensures that mediated institutional access for academics and student users is accurate, comprehensive, and timely.

2.1.4 Existing Endorsers Migrating To Phase II

Content providers who previously endorsed the Phase I guidelines will need to update their endorsement for Phase II compliance. This will involve re-submitting sample files for testing by the knowledge base vendor members of the group. The timescale for compliance with Phase II recommendations will be six months following publication of the final Phase II Recommended Practice.

2.2 Collaboration Activities

The KBART initiative and recommendations do not exist in isolation; they rely on community involvement and take-up to improve the landscape for the end user and by doing so make the supply chain more efficient and effective. As described in section 1.2, the project has used data from surveys as well as feedback from speaking engagements, outreach events, and mailing lists to inform the guidelines.

There are a number of bodies and initiatives that are working in the area of supply chain metadata and discoverability improvement with whom KBART is liaising to improve collaboration and metadata provision after publication of the guidelines. In other words, this is a constantly shifting landscape. This section covers the organizations and initiatives that KBART is liaising with and possible future outcomes that will further improve this area.

2.2.1 EDItEUR and ONIX for KBART

The value of the KBART format (a tab delimited .txt file) is that it can be used for both human-readable and machine-readable purposes and can be supplied on a webpage and/or transmitted via FTP between

content provider and knowledge base vendor systems. It is also a low cost approach to effective knowledge base metadata transfer for publishers who do not have substantial in-house development expertise or resources.

For content providers who value the potential of automated upload and use of an extensible format for KBART metadata communication, the KBART group is discussing with EDItEUR the implementation of an ONIX for KBART message, potentially based on the existing ONIX-SOH message. It may be of particular value to content providers who are already using the ONIX suite of messages for other business requirements.

As part of such a format, a KBART to ONIX mapping would be developed to ensure that metadata requirements match, both in terms of structure of the message and purpose. There are ongoing discussions regarding endorsement and sustainability as well as the message itself, but this is a positive step for the transfer of knowledge base metadata.

2.2.2 Shared Knowledge Base Initiatives

Since the publication of the Phase I guidelines, there has been a gathering of momentum towards community-owned central knowledge bases. These shared services offer tremendous potential in terms of community verification of content provider metadata and in making that metadata available to knowledge base vendors.

Currently the emphasis is on the many-to-many transmission of metadata from content providers to multiple knowledge base vendors, which is inefficient and relies on resources within sender and recipient organizations to develop and process files. In addition, when a library knows that metadata is incorrect, its correction can often be overridden later by the same incorrect content provider metadata. The shared service approach offers potential to verify content within files and provide that assurance to knowledge base vendors on behalf of the community.

Examples of current initiatives that are making KBART files available to knowledge base vendors include:

- **KBART title lists** – Endorsed content providers are making their title lists in KBART-compliant format available to the community. Links to these lists are provided in the KBART Registry.
- **Knowledge Base+** (www.kbplus.ac.uk/kbplus/) – Run by Jisc Collections, KB+ is a shared community service knowledge base for UK academic libraries to support the management of electronic resources. It provides accurate and up-to-date data, such as publication, licensing, and subscription information. KB+ is KBART compliant. All KB+ KBART files are available at www.kbplus.ac.uk/kbplus/publicExport. KB+ data is used by institutions and link resolver providers such as EBSCO, Ex Libris, OCLC, and ProQuest (Serials Solutions).
- **The Global Open Knowledgebase (GOKb)** (gokb.org/) – An international collaborative project led by Quali OLE and Jisc, GOKb is an open, community-managed knowledge base to support management of electronic collections in libraries. GOKb contains globally-true data about packages, titles, platforms, organizations, and licenses.

2.2.3 Licensing

The KBART Working Group has been collaborating with library consortia to provide guidance on statements within licenses and contracts relating to KBART compliance. This is an important part of community adoption of the Recommended Practice and has led to model license inclusions that are then negotiated with content providers as part of contract renewals.

Section 3: Metadata for Consortia

3.1 Introduction

It is common for libraries to form a consortium to better negotiate purchases with content providers. Often these consortia will purchase unique packages of titles available only to their member libraries. Although not available to every library, these packages serve the needs of multiple institutions. Libraries and their consortia staff need to track the titles available to them, their dates of available coverage, URLs for title access, and other details in the same way that they do for packages purchased outside of the consortium structure.

Historically, it has been difficult for libraries and linking vendors to obtain accurate title list data from content providers for these non-standard packages. Some content providers have only been able to provide minimal title details for consortia packages (e.g., title and ISSN) such as found in marketing lists, while others have not been able to provide consortia package data at all. Consortia representatives feel that accurate title lists for consortia packages are important, and the KBART Working Group recommends that content providers work to establish feeds for consortia-specific packages when applicable.

3.2 Consortia Survey Summary

As discussed in section [1.2](#), The KBART Working Group conducted a web survey in January 2012 to address the importance of KBART deliverables to consortia and the libraries they serve. About 20 of the respondents represented consortia that serve about 2000 libraries worldwide, in the aggregate. The remaining 180 respondents represented individual libraries served by those consortia. The full results of the survey (including anonymous individual comments on many questions) are included as [Appendix B](#).

The consortia representative respondents:

- agreed (nearly all) that accurate “standardized” holding lists were very important to members of their consortium (Question 17);
- felt it was very or somewhat important for the KBART Working Group to address holdings list generation distribution and accuracy at the consortial level (Question 18);
- said that they were either only somewhat effective or not very effective in providing these lists (Question 19);
- felt that consortia should be concerned with distributing them to a significant extent (Question 20); and
- indicated that addressing open access content, establishing a clearinghouse of downloadable standard holding lists, and adding e-journal publisher endorsers were their top three priorities for the KBART initiative (Question 21).

Library representative respondents:

- indicated that standardized lists were very important (Question 29);
- perceive their knowledge base accuracy to be about 80% (Question 32) though they strive for greater than 95% accuracy (Question 33);

- ascribed high importance to those lists that pertain to large group of libraries (Question 34), those lists that pertain only to multiple members of their consortium (Question 35), and those that pertain only to their individual library (Question 36);
- update their lists annually (Question 37) though more than half said they would prefer monthly or even weekly updates (Question 39); and
- clearly prioritized the population of a clearinghouse of downloadable standard holdings lists and addressing open access content as the two most important KBART initiatives (Question 40).

The results of this survey were used to inform the following recommendations.

3.3 Recommendations for Consortium-Specific Files

A consortium-specific file (or files) should be created if:

- The package (or packages) has a customized set of titles that are available for purchase by a particular consortium.

Content providers that sell packages of titles that are available globally should not provide consortium-specific files for those same packages. Use the generic title list instead.

- A package contains customized dates of coverage.

For example, if a content provider normally provides access from 1995 to present, but sells a consortium a package with dates from 2000 to present, a consortium-specific list should be created with the titles the consortium has purchased with their custom coverage dates.

3.4 File Contents and Naming Changes

3.4.1 File Name

The file should be entitled:

[ProviderName]_[Region/Consortium]_[PackageName]_[YYYY-MM-DD].txt

“Region/Consortium” should include any information based on where the package is sold, or to what consortium it applies. If the file is for a universal list, “Global” should be used.

Publishers and vendors should avoid using special characters in filenames

Examples:

Title list is not region or consortium-specific; it includes all titles from the content provider:

- JSTOR_Global_AllTitles_2013-01-14.txt
- TaylorandFrancis_Global_AllTitles_2012-08-30.txt

Title list is consortium-specific for a specific package:

- IOP_NESLi2_Option1-2011_2012-05-31.txt
(includes a year as part of the package name)
- OxfordJournals_SCELC_AllTitles_2013-01-09.txt
(contains all titles that the consortium has subscribed to)

Title list is region-specific, for a specific package:

- Springer_Asia-Pacific_Medicine_2013-01-28.txt

3.4.2 When to Produce Separate Files

Separate files should be produced for each package of content that the provider offers. Files should be named as customers would expect to see them labeled in the knowledge base, using the syntax:

[ProviderName]_[Region/Consortium]_[PackageName]_[YYYY-MM-DD].txt

Example:

JSTOR_Global_ArtsandSciencesV_2013-01-14.txt

Providers and recipients can agree in advance how best to present complex package (collection) names.

In conclusion, the changes to support data communication for library consortia using the KBART recommended practices center entirely upon conditions where a consortium-specific file is required. No additional fields or changes to the title-level metadata are required, beyond what has already been stated by the KBART Working Group in the Phase I and in these Phase II recommendations.

...

Section 4: Open Access Metadata

4.1 It's Free Stuff Isn't It? Why Is Metadata for OA Required?

Users can access free and open access content through non-library specific search engines, in institutional repositories, and directly on publisher websites at article level. So it is worth re-stating why there is a need for content providers to send metadata to link resolvers for free and open access content. This leads us back to the importance of link resolver technology. Its value lies in aggregating e-content delivery appropriate to the user affiliation in a single place and delivering appropriate, authenticated links to content that cannot be delivered through a generic search engine. As users interact with library technology (such as link resolvers and resource discovery services), they are presented with all of the content that their home institution licenses on their behalf. For the library and their users it therefore makes sense that they can link to all the content delivered by a publisher together—both free and fee-based. For the publisher it means that traffic coming through library technology tools is presented with the entire content portfolio, not just that which is paid for by subscription. This is increasingly important in an environment of evolving business models for scholarly information dissemination.

Open Access electronic content holdings have been problematic both for librarians and for users for a variety of reasons that can differ from those of traditional fee-based electronic content. Since the release of the KBART Phase I Recommended Practice, Open Access content has become more and more prominent. Many large publishers offer authors the ability to publish their content as an Open Access article either in a fee-based journal (hybrid OA) or in a 100% OA journal. There are Open Access books being published as well. In short, this is a growing area and a mode of electronic content access that cannot be overlooked by library technology vendors, libraries, or content providers keen on improving content discovery. KBART set out in Phase II of the Recommended Practice to investigate problems surrounding metadata for Open Access and free content in order to improve access to such content.

Holdings metadata for Open Access and freely available content poses a number of unique challenges. First and foremost, it should be stated that there are numerous definitions of Open Access from libraries, publishers, and vendors alike. Because there are so many definitions of Open Access and because it is essentially an evolving definition, KBART has made the conscientious decision not to differentiate between these varying types of open access content. Additionally, KBART recognizes the fact that there is indeed a difference between Open Access and freely available content, but in the end, users simply want access and are not necessarily concerned with business models. KBART facilitates improved access and holdings metadata transmission regardless of business model and is treating Open Access content no differently from freely available content on publisher platforms.

4.2 Open Access Survey Summary

In the consortia and open access metadata survey (see section [1.2](#) and [Appendix B](#)), the KBART group asked a number of questions about Open Access. The answers reflect the complexity and also the importance of this area. Nearly 70% of respondents indicated that it would be useful to identify Open Access and free content at both the journal level and at the article level. When asked if librarians would like to see Open Access and free content grouped together as a package or interspersed amongst vendor and publisher packages, users were mostly split; 55% indicated that they would like to see it grouped together while 45% preferred it to be interspersed. The vast majority of respondents considered indications on the user interface that a journal or article is Open Access or free as either very or somewhat important. In addition, many respondents added comments raising further issues. For example, many respondents noted that they (and their users) struggle with hybrid Open Access content. Additionally, many respondents expressed concern about determining whether or not there are content embargoes on

hybrid Open Access titles and substantial frustration when looking at temporarily free content that may disappear at an undetermined time. There was also a clear interest in supplying Open Access and free content metadata at the article level. In this phase KBART cannot address all of the issues raised in the survey and some are out of the KBART scope. We do however recommend that Open Access and free content metadata issues be dealt with and we will carry out further work on Open Access material in the future.

4.3 Consideration of Types of Open Access

Often it is not easily determined whether an article or title is available via Open Access (OA), free, or by fee. There are numerous possibilities in this regard. Hybrid OA titles represent the biggest problem with regard to OA electronic holdings metadata. A single or numerous article(s) within a journal title may be available via Open Access supported by author fees while the rest of the journal is only available via subscription. In some cases, a publisher will choose to make everything Open Access after a period of time has elapsed since a volume or issue has been published. Consider the following Open Access scenarios:

- **Delayed OA** – a title that has all but the most current 12 months of a journal title available via Open Access but the most recent 12 months of content requires a subscription
- **Title Transfer OA** – a title which changes from Open Access or freely available to one that requires a subscription and/or back again when bought and sold by various publishers
- **Author Pays Hybrid OA** – a title that, by and large, requires a subscription for content access, but certain articles are available via OA supported by author fees

In all of these cases, it is very difficult to ascertain why one may be denied access at any given time. Because a KBART metadata holdings file is at the title level, and not at the article level, it is nearly impossible to account for the wide variety of Open Access and freely available content models. In the future, KBART recommend that this indication also be dealt with at the article level for best results. Article-level communications are not included in the current KBART scope of work.

4.4 Changes in the Recommendations

For Phase II of the Recommended Practice, KBART has changed one field (see [Table 1](#)) and added an additional field and optional free text metadata (see [Table 2](#)) that will hopefully improve access for end users and make trouble-shooting access problems easier for libraries and content providers.

Table 1: Changed field in Phase II

Field Name	Description
notes	For content providers who wish to include notes about Open Access coverage, particularly Hybrid Open Access titles, KBART will employ the use of the optional “Notes” field. “Notes” will remain a free-text and optional field in KBART Phase II and is the easiest place to explain subtle nuances about a particular title (including, but not limited to, notes about Open Access content) that may be useful for third party vendors, such as link resolver suppliers, as well as libraries.

Table 2: Additional required field in Phase II

Field Name	Description
access_type	This newly required field can have two values. • F – Use this value to indicate that a title is Open Access or has free content. • P– Use this value to indicate that a title requires payment of a fee of any kind (e.g., subscription, purchase, etc.) for content access. In the case of a hybrid title with both free and paid content, publishers and vendors should repeat title records as many times as necessary to indicate coverage ranges for which access is free and coverage ranges for which access is paid. For example, if a journal has a one year moving wall where the most recent 12 months are paid access, this would be indicated by one record having a value of “P” (for paid) in the access_type field, a second record showing a value of “F” (for free) in the access_type field for the content older than 12 months. NOTE 1: Hybrid Open Access titles are problematic because KBART cannot yet handle article-level holdings metadata. If a title has some Open Access or free content, the access type for a title should be indicated as fee-based, with the value “P”. For a title to use the “F” value, 100% of the content must be free. NOTE 2: When combining access_type and embargo_info, KBART recommends repeating the holdings information for such a title twice. One record would be for the free content ahead of the moving wall (coded as “F” for free) and the second record would be for the content that requires a subscription (coded “P” for paid). Of course, the appropriate access dates would go in their respective records.

Section 5: E-Books and Conference Proceedings

5.1 Importance of Metadata

The importance of metadata for e-books has been apparent to the publishing, library, and e-commerce communities for some time. Bibliographic metadata relating to a specific book is still key and bibliographic metadata standards have evolved to take account of e-book formats.

However, in the e-book era, the traditional catalog record (supplemented by URLs and content provider information) is not the only or best way of linking to the e-book on the appropriate content provider's website(s). Webscale resource discovery services (utilizing link resolvers and library catalogs as underlying data sources) can ingest bibliographic records (e.g., MARC 21 records). However, the catalog record is not dynamic enough as a record of electronic holdings when holdings can represent multiple platforms with different access restrictions, and where these platforms may come and go. It is the link resolver (and knowledge base) that provides for the link from discovery tool to content—linking seamlessly to both e-journal and e-book content from one place. The link resolver knowledge base can be more current than the traditional library catalog because the library catalog requires the library to source and import MARC records and updates from the content provider on an ongoing basis. The link resolver knowledge base provider can set up, with minimal effort by the library, an ongoing data feed from the content provider with information to enable the discovery and access to e-book content. Webscale discovery tools are increasingly merging the bibliographic elements from the MARC record with the electronic coverage statement, allowing both discovery and linking from the knowledge base to full-text content.

The library community is, therefore, looking to the content provider to provide accurate, timely, and comprehensive e-book electronic holdings metadata to link resolver knowledge bases to facilitate ease of discovery through library discovery tools and to increase efficiency of administration by the library. So far, quality of e-book metadata delivered by content providers has proven to be problematic. Important metadata can be missing (e.g., ISBN of the electronic version), out-of-date, or simply incorrect. Moreover, each content provider, restricted by its own production systems, sends its e-book metadata in its own unique format. This requires link resolver vendors to develop proprietary import rules for individual content providers, which leads to delays in import and inefficiency. Finally, since a content provider will produce new e-book titles much more frequently than new e-journal titles, the accuracy and timeliness of the e-book metadata becomes even more important to ensure that new content is discoverable immediately on publication.

Standardization of metadata for conference proceedings needs special attention. Since conference proceedings are a hybrid of a serial and a monograph, the metadata should include information for the serial title as well as information on the volume level, and a connection needs to be made between the serial and the volume. The connection between a conference proceeding volume and its parent serial title is important because many conference proceedings include hundreds, sometimes even thousands, of volumes and span decades. Without this connection, these volumes cannot be grouped together.

Content providers, resellers, and librarians all understand the importance of establishing an accurate, comprehensive, and up-to-date knowledge base for e-books. The standardization of e-book metadata in the knowledge base has become more and more of a priority as the take up of e-books escalates. Although the KBART Phase I Recommended Practice was published in 2010 and some of the recommended fields may be applied for e-books, some special clarification for the use of these fields for e-books is needed and some additional fields are needed to purely support e-books and conference proceedings scenarios.

5.2 Changes of Title

Another challenge that the KBART Phase II recommendations addresses is how to deal with the problem of changing titles for serials—both journals and conference proceedings. Many journals and conference proceedings change title during their lifespan, sometimes more than once. The Phase I Recommended Practice does not include a field that connects the various titles over the history of the publication (where it is still represented online). Many librarians have commented on the difficulty of finding and maintaining the relationships between the old and the new serial titles where linking relationships need to be made between former and current titles. This topic is described further in the NISO Recommended Practice, *PIE-J: Presentation & Identification of E-Journals* (NISO RP-16-2013).

KBART Phase II recommends a new field, `preceding_publication_title_id`, to establish this historical connection for serials, both journal serials and conference proceeding serials as described in [Table 3](#).

Table 3: New field for title changes in Phase II

Field Name	Description
<code>preceding_publication_title_id</code>	The title identifier of any preceding publication title for serials and conference proceeding serials. The publisher's proprietary identifier (as described in section 6.6.13) should be used where it exists. Examples: 40026030 0146-6453

5.3 Changes in the Recommendations

There are two main sets of changes from the Phase I RP that are described in this section.

First, annotations are given to several existing metadata fields (in section [6.6](#)) to differentiate their use in monographs vs. serials.

- Some fields, specified in the Phase I RP are specifically for e-journals and do not apply for e-books. These are:
 - `date_first_issue_online`
 - `num_first_vol_online`
 - `num_first_issue_online`
 - `date_last_issue_online`
 - `num_last_vol_online`
 - `num_last_issue_online`
- Some fields can be used for both e-journals and e-books (including conference proceedings), but in different ways. For instance, `print_identifier` and `online_identifier` for e-journals refer to ISSN and eISSN. However, the same fields for e-books should mean ISBN and eISBN.

- For e-journals, `publication_title`, `title_url`, and `title_id` are clearly for metadata on the serial level. For e-books, the same fields are clearly for metadata on the monograph level. However, for conference proceedings, multiple rows to the metadata file will need to be added. On the serial level, `publication_title`, `title_url`, and `title_id` refer to the title, URL, and identifier of the conference proceeding serial. On the monograph level, these fields apply to the specific volume or book within the conference proceeding series.
- One field from the Phase I Recommended Practice, `first_author`, should now be used for monographs only.

Second, seven new metadata fields are recommended for monographs and one new field is recommended for serials, including both journal series and conference proceedings series.

- A new field, `publication_type`, is introduced to differentiate serial (i.e., journals and conference proceeding series) from monograph (i.e., book, e-book, and conference proceeding volume).
- Four new monograph-specific fields are recommended:
 - `date_monograph_published_print`
 - `date_monograph_published_online`
 - `monograph_volume`
 - `monograph_edition`
- For both e-books and conference proceedings, the new field `first_editor` is introduced, to be used along with the existing field `first_author`.
- To indicate the relationship between a conference proceeding volume and the conference proceeding series, the field `parent_publication_id` is recommended.
- The new field `preceding_publication_title_id` is recommended to establish the relationship between a successor serial title and its preceding serial title.

These new fields are summarized in [Table 4](#).

Table 4: Summary of new e-book and conference proceedings fields

Field Name	Description
<code>publication_type</code>	Includes the value “serial” (i.e., journals and conference proceeding series) or “monograph” (i.e., book, e-book, conference proceeding volume).
<code>date_monograph_published_print</code>	Date the monograph is first published in print.
<code>date_monograph_published_online</code>	Date the monograph is first published online.
<code>monograph_volume</code>	Number of the volume of a monograph (applicable to e-books and conference proceedings; for proceedings, volume within the conference proceedings series).
<code>monograph_edition</code>	Edition for a book.

Field Name	Description
first_editor	First editor (for monographs, i.e., e-books or conference proceedings volumes).
parent_publication_title_id	Title identifier of the parent publication. For a conference proceeding volume, the <code>parent_publication_title_id</code> is the <code>title_id</code> of the conference proceedings series.
preceding_publication_title_id	For serials and conference proceeding serials, the title identifier of any preceding publication title.

[Section 6](#) summarizes all KBART Phase II Recommended Practices and should be used for preparing systems and sample files for KBART Phase II endorsement and compliance.

Section 6: Guidelines for Effective Exchange of Metadata with Knowledge Bases

6.1 Method of Exchange

Content providers should post holdings data to a dedicated metadata webpage or FTP (File Transfer Protocol) site for download by link resolver suppliers. This minimizes the effort involved in the transaction for both parties. FTP is a simple protocol for allowing users to exchange files across a network. 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 interests and, therefore, 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 an FTP site is unachievable, then e-mail is an acceptable alternative. The data from the tab-delimited file (see section [6.4.1](#)) 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 [6.5.1](#).

6.2 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 should update data more frequently than once a month if their holdings change more frequently. This will particularly be the case with e-books for current publication year works. Updates should be posted to the metadata webpage or FTP site with an appropriate date stamp as indicated in the filename conventions (see section [6.5](#)).

6.3 Data Contacts

Both the content provider and the knowledge base developer should designate specific staff members to be responsible for metadata 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 has provided the KBART Registry (<https://sites.google.com/site/kbartregistry/>), which includes a web form for registering the organizational contact, as well as the platform and metadata URL links.

6.4 Data Format

6.4.1 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 while they can be "commented out," doing so leaves a greater opportunity for error than the