Six Impossible Things: Moving KBART into the Next Decade

Andrée Rathemacher
University of Rhode Island, andree@uri.edu

Robert Heaton
Utah State University, robert.heaton@usu.edu

Noah Levin
NISO KBART Standing Committee, noah.levin626@gmail.com

Christine Stohn
Ex Libris, Christine.Stohn@exlibrisgroup.com

Andrée Rathemacher, Robert Heaton, Noah Levin, and Christine Stohn, "Six Impossible Things: Moving KBART into the Next Decade" (2019). *Proceedings of the Charleston Library Conference.*
http://dx.doi.org/10.5703/1288284317173

Six Impossible Things: Moving KBART Into the Next Decade

Andrée Rathemacher, University of Rhode Island, andree@uri.edu

Robert Heaton, Utah State University, robert.heaton@usu.edu

Noah Levin, NISO KBART Standing Committee, noah.levin626@gmail.com

Christine Stohn, Ex Libris, christine.stohn@exlibrisgroup.com

Abstract

KBART is one of the most successful NISO recommendations today. Formally supported by over 80 organizations across all stakeholder groups, it enables a standardized transfer of data between content providers and knowledge bases. Most recently KBART added an automated process to transfer holdings data to localize an institution's knowledge base holdings. While KBART was originally built to focus on journal and book data, the world has moved on—the different flavors and nuances of open access, the increased use of audiovisual material, holdings at the chapter and article levels, and issues around translations, transliterations, and author names are just some of the challenges that are disrupting the flow. So what is next for KBART? How does it adapt to continue to solve the data flow problems that libraries, publishers, and knowledge base providers face today? The presenters in this session, all members of the NISO KBART Standing Committee and/or the KBART Automation Working Group, discuss the status and future of a "Phase III" revision of NISO KBART that aims not only to clarify the existing recommendations but also to expand them to address the new challenges, including the support of additional content types beyond serials and monographs and improvements to item-level discovery and access.

KBART: A Short Overview

KBART (Knowledge Bases and Related Tools) recommends best practices for the communication of electronic resource title lists and coverage data from content providers to knowledge base developers. KBART specifies file format, delivery mechanisms, and fields to include, and it applies to both serials and monographs. Knowledge bases are used to provide data for OpenURL link resolvers and to populate library discovery systems with an institution's e-resource holdings data. Many libraries also use knowledge base data in library catalogs, for e-journal title lists, in electronic resource management systems (ERMSs), and in other tools. If a knowledge base contains inaccurate information or is not updated regularly, these discovery tools will fail. By providing a recommended practice for communicating information from content providers to knowledge base developers, KBART helps ensure the integrity and functionality of knowledge bases. The Phase I KBART Recommendation was published by NISO in 2010, and in 2014 KBART Phase II extended Phase I recommendations, specifically with regard to consortia packages, e-books, and open access content.

In 2019, KBART Automation was released. This is a companion NISO Recommended Practice (RP) to KBART that provides for the automatic transfer via API from content providers to knowledge bases of institution-specific KBART-formatted holdings files.

In short, the goal of KBART is to increase the accuracy of knowledge base content to reflect accurate title list and package/collection offerings of content providers. KBART Automation enables the automatic setting of local holdings in knowledge bases by transferring KBART-formatted institutional holdings files from content providers to knowledge bases.

Changes in the Information Landscape and KBART Phase III

There have been a number of changes in the information landscape since KBART Phase II was released that the KBART Standing Committee seeks to address during the drafting of KBART Phase III. One is the increasing granularity of the level at which access to content is determined, for example access to journal content at the article level due to hybrid open access and to book content at the chapter level due to changing content provider sales models. There is more content available than ever before; items that used to number in the hundreds of thousands now number in the millions. Many more material types are available beyond journals and e-books, such as book chapters, audio material, images, films, manuscripts, and maps. Content providers continue to

https://doi.org/10.5703/1288284317173

develop new business models, and KBART needs to be able to adapt to accommodate them.

Currently, the KBART Standing Committee is in the process of developing a proposal for KBART Phase III for approval by the NISO Information Discovery & Interchange Topic Committee. Once the proposal is approved, we will develop working areas and subgroups as necessary. We will identify areas of expertise needed and recruit new members. The subgroups will research possible changes—with emphasis on current content provider practice and what knowledge bases can utilize—and then create an outline of proposed new recommendations. A draft of KBART Phase III will be circulated for a 30-day comment period, during which time the KBART Standing Committee will engage in marketing and education. Public comments will be incorporated into the document before final publication.

Low-Hanging Fruit

In the more than five years since KBART Phase II was released, a number of additional needs have been identified that the Recommended Practice should address. Some of these needs are relatively straightforward. In addition, the KBART Standing Committee has learned that certain areas of the KBART Recommended Practice would benefit from additional clarification, even if they remain unchanged. We're calling these straightforward updates and clarifications "low-hanging fruit."

In Phase III, we plan to include more guidance and examples for most requirements, since content providers that are new to KBART sometimes struggle to get started with bringing their files into compliance. We will offer expanded guidance on what files to create and what metadata to include, for example whether to create separate "All Titles" files for serials and monographs and expanded criteria on when to create a new file for a package of content. We will also provide clarifications and additional information on each data field as well as more examples of correct implementation. Examples include alternative ways to represent title histories for providers who might not have unique identifiers for previous titles; how to handle a combined volume or issue number, for example 3/4; how to represent issues that are supplements having a different title from the mother publication but sharing an ISSN; and whether to require an end date for journals that have not ceased but are significantly behind in publication.

Many content providers have an extensive catalog of content for sale by content type, subject, geographic region, consortium, etc. Per the KBART RP, this results in a separate KBART file for each offering. The number of files can in some cases make it difficult to keep track of what has been added, removed, or changed. For Phase III, we are considering recommending that content providers create a document that serves as a guide to their KBART files. A manifest file could include the names of the files delivered, the collection name that each file represents, a unique identifier for each collection, a description of the collection, the number of records in the collection, and the date the KBART file was created.

A related issue is tracking content withdrawn from packages. Libraries that purchased content that is subsequently withdrawn often retain grandfathered access to that content. However, since KBART files do not usually include content no longer available for purchase, such content becomes invisible to link resolvers and disappears from discovery systems in those libraries that retain access. KBART Automation avoids this problem by relying on library-specific holdings files. KBART Phase III could address this issue by requiring a version history of files or the addition of add-delete-delta files to flag changes. Knowledge base vendors could then develop solutions around these files that would enable libraries to continue to manage their access to withdrawn content in the knowledge base.

KBART Phase II only addresses holdings data for serials and monographs. Since the last RP revision, there has been an increase in the popularity of textual content that falls outside of these categories, for example blogs, transcripts, manuscripts, and data sets. In addition, there has been a growth in nontextual content such as audio, video, and images. Content providers have been forced either to add a field to the end of their KBART files to identify such content or to exclude the content from their KBART files. At best this causes confusion; at worst valuable data is excluded from KBART holdings. In Phase III, we will develop support for these additional content types.

Just as content types have expanded since KBART Phase II, so has the prevalence of global content, as KBART has been more widely adopted around the world. Currently, KBART files do not allow for the identification of translations or the representation of author names or titles in multiple languages. To improve support for global content in Phase III, we

hope to be able to support content with non-Latin characters, translated titles, transliterated titles, and language of content. We will also look at expanding the "first_author" and "first_editor" fields to include full names, which would better identify content in languages that have very common surnames.

In KBART Phase III, we are also looking to overhaul the endorsement process and to establish varying levels of endorsement. In this way we can reward content providers who meet all of the recommendations in the RP while also making endorsement easier for content providers unable to attain 100% compliance. Additionally, a tiered endorsement structure will allow us to distinguish between content providers that have achieved Phase III versus Phase II compliance. We also plan to better define what KBART compliance means when applied to knowledge base vendors.

Finally, in KBART Phase III we intend to include model license language for libraries that would like to include such language in their contracts with content providers. A number of current model licenses mention KBART files, such as those from California Digital Library (2016), LIBLICENSE (2015), CKRN (2016), and Jisc (2018). By providing a standard license clause requiring that content providers provide holdings files in KBART-compliant format, we hope to make the adoption of such contract language easier and to encourage the spread of KBART.

Tough Questions

In their preparations for KBART Phase III, the KBART Standing Committee recognized a number of changes in the information marketplace as well as new needs of library and content providers that are not as easily addressed by KBART as it exists. We have named these issues—and how and whether KBART can provide a solution—"tough questions." At stake is the appropriate scope and purpose of KBART as it is used today and a balance between the granularity of metadata communicated in KBART files versus maintaining the Recommended Practice's simplicity.

As a first step, the KBART Standing Committee feels that it is important in KBART Phase III to acknowledge the current uses of KBART files. As noted, KBART originated in 2010 as a recommendation for providing standardized data to OpenURL link resolvers in support of reliable citation-to-full-text linking. Now, in part due to the wide adoption of KBART, KBART files are being used in ways not anticipated when the KBART RP was first drafted almost a decade ago. For example, KBART files, through knowledge bases, are used to display and link to library holdings in discovery systems and A–Z lists. Librarians use KBART data to populate their ERMSs in order to track what they purchase along with associated costs and usage. KBART files are also used to conduct overlap analysis between print and electronic holdings and to compare publisher packages when considering purchases.

With KBART Automation, identifying holdings at the institutional level became a central focus of KBART, which was an advance since library holdings do not always correspond to the "one size fits all" packages presented in knowledge bases. KBART Automation brought new needs and possibilities that will affect the core purpose of KBART. Thus, in the Phase III revision, we want to be sure to acknowledge the role and importance of KBART in today's e-resource ecosystem and revise the mission of KBART to reflect how it is currently being used.

Perhaps the biggest "tough question" is the issue of article- and chapter-level metadata. As noted above, the growth of hybrid open access in journals results in access determined at the article level as opposed to the journal title level. Archival primary source databases incorporate content that is extremely granular. New business models are in development that will result in publishers selling topical packages that include individual articles and chapters. Since KBART communicates coverage at the title level only, it cannot accurately represent such granular access. In the case of access determined at the article or chapter level, access communicated by KBART would necessarily be incorrect.

The questions, then, for the KBART Standing Committee and our user community, are should article- and chapter-level access be communicated using KBART, and how could it be done? Is there another way to communicate article- and chapter-level access that would tie into KBART but exist outside of it? One concern is that current knowledge bases are not set up to support article- and chapter-level data, so even if KBART files included this data, knowledge bases could not ingest them. An alternative solution to KBART for communicating access at a granular level might be through an API that requests and communicates access information on the fly by consulting a database. For example, Unpaywall.org is a database that includes information on open access articles and can be queried via API.

Another "tough question" is whether KBART should support the XML file format. Currently, KBART requires tab-delimited TXT files, which work well for both knowledge bases and users who wish to reuse them in spreadsheets and other off-the-shelf tracking tools. During the drafting of the KBART Automation RP, discussion of support for XML files in the KBART RP was raised. If KBART optionally allowed for files to be created in XML in addition to TXT, this would give content providers additional flexibility. XML is preferable for some developers and is the supported data exchange format for some providers' systems. In addition, XML can better accommodate larger files and would allow future support for communication of more complicated data. On the other hand, knowledge base vendors currently support TXT files and current processes are working well. To create the option for content providers to deliver KBART files in XML would require that knowledge base vendors create the capability to ingest XML files in addition to TXT files, which might be a burden. Because KBART has been so widely adopted, we need to be cautious that we do not make changes that interrupt current workflows or potentially negatively impact the adoption of KBART.

Open Discussion

The presenters sought input on these and other questions from attendees. One librarian asked if KBART could support links at the e-book title level that linked directly to PDF and EPUB downloads. The presenters responded that while knowledge bases do not typically use the links in KBART files for linking (they use links constructed by the knowledge base software), this was worth considering, and that it might be possible to add additional columns to the KBART file to accommodate direct links to content.

A representative from a content provider noted that subscribing library consortia often requested from them additional information, including title-level DOIs, frequencies, and subjects for journals. Because they do not want to create separate title list files, the content provider adds this information to additional columns in their KBART files. It would be nice if KBART included a field for DOI or provided guidance in including DOIs. Presenters responded that a DOI does not always link to content on the desired platform, as the same content can exist in a variety of places. Further, KBART is only intended to communicate the metadata needed to identify an item in a knowledge base; anything more would be beyond its scope and purpose. KBART's simplicity is part of its success story. Presenters agreed they would bring the question of including a field for DOI at the title level to the KBART Standing Committee to consider for inclusion in Phase III.

Attendees were encouraged to fill out an online survey on their priorities for the KBART Phase III revision.