KBART Phase III Update

Andrée Rathemacher
KBART Standing Committee Co-Chair
Nettie Lagace
Associate Executive Director, NISO

2020 NETSL Annual Spring Conference
Friday, April 17, 2020

What is KBART?

Knowledge **B**ases **A**nd **R**elated **T**ools

UKSG/NISO Working group (2007-2014)

NISO Standing Committee (2014-present)

What is KBART?

Recommendations for the transmission of data to knowledgebases (KBs)

Title and package data of electronic serials and books

Purchased/leased/accessed by libraries

Transparency of data exchanged

Who is the KBART Standing Committee?

Co-chairs

Noah Levin
KBART Standing Committee Co-chair

Andrée Rathemacher
Professor / Head, Acquisitions
University of Rhode Island

Members

Davin Baragiotta
IT Lead
Consortium Érudit

Dominic Benson
Analytics and Discovery Officer
Brunel University London

Courtney Bremer
eResources Access Manager
University of Waterloo Library

Mark Calkins
Senior Product Support Specialist
OCLC

Stephanie Doellinger
OCLC Online Computer Library Center

Elif Eryilmaz-Sigwarth
Technical Product Manager
Springer Nature

Jason Friedman
Manager, Member and Metadata Services
The Canadian Research Knowledge Network

Robert Heaton
Collection Management Librarian
Utah State University Libraries

Benjamin Johnson
Engagement Manager, Provider Relations
ProQuest

Sheri Meares
EBSCO Information Services

Jonathan Ponder
ITHAKA/JSTOR/Portico

Christine Stohn
Senior Product Manager, Discovery & Delivery
Ex Libris/Proquest

4
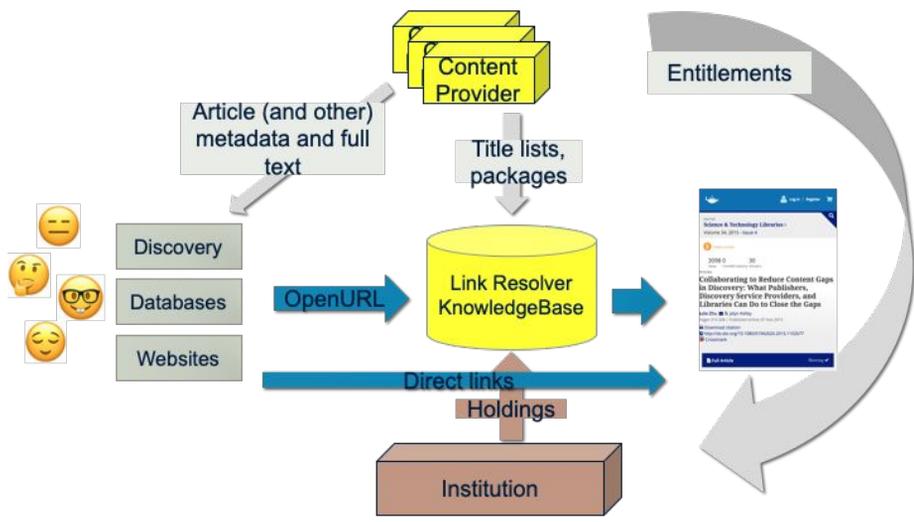
Why KBART?

- Culling, James. "Link Resolvers and the Metadata Supply Chain" (2007) (https://www.uksg.org/sites/uksg.org/files/uksg_link_resolvers_final_report.pdf)
- Inconsistent holding list metadata format
  - Embargo period format
    - Example relative vs. absolute embargo?
  - Date/enumeration formats
    - MM-DD-YYYY? / DD-MM-YYYY?
- Inconsistent metadata update procedures ...

Discovery/Delivery Landscape

KB management before KBART (libraries)

Proactive reconciliation of an ejournal package list:

- Request title list with detailed holdings info from publisher (repeatedly, naggingly)
- Compare with that of your subscription agent and KB vendor
- Now that you have 3 (or more) different title lists, translation phase includes dealing with:
  - Number of titles and titles themselves
  - ISSN mis-matches
  - Title changes, mergers, acquisitions, new starts, and losses
  - Publisher-reuse of ISSNs/title combinations
  - Reconciling date discrepancies manually (and inconsistent/unclear formats)
- Go live
- Lather, rinse, repeat!

KBart

NISO
How the information world
CONNECTS

KB management before KBART (KB vendors)

- Similar to the library's process
- Need to contact providers again and again
- Invest a lot of time correcting data problems
  - Investigating end-user queries and complaints
- Update procedures vary by provider
- If unable to get data from provider, may resort less preferable acquisition methods (web site inspection)
  - Last resort, not preferred

KBart

NISO
How the information world
CONNECTS

KBART Sample

Note: The spreadsheet for this data has been split into three sections for viewability in this document. The second and third sections represent additional columns that were appended to the first section in the original spreadsheet.
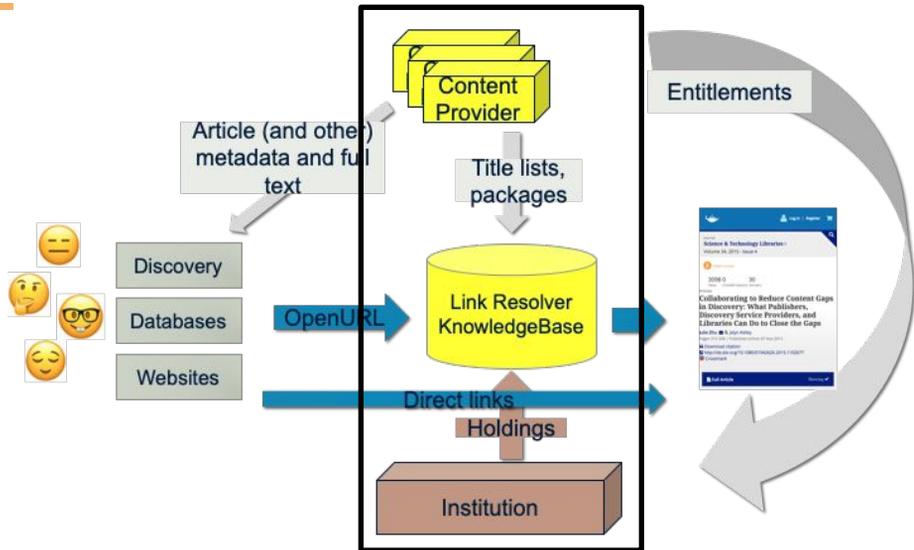
publication_title	print_identifier	online_identifier	date_first_issue _online	num_first_vol _online	num_first_issue _online	date_last_issue _online	num_last_vol _online	num_last_issue _online
Dante and Aquinas	978-1-909188-03-7	978-1-909188-07-5	2013-05-15					

title_url	first_author	title_id	embargo _info	coverage_depth	notes	publisher_name	publication_type
http://www.ubiquitypress.com/files/DanteAnd Aquinas.pdf	Ryan	10.5334/bad		fulltext	Published under CC-BY license	Ubiquity Press	monograph

date monograph published _print	date monograph published _online	monograph _volume	monograph _edition	first_editor	parent publication title_id	preceding publication _title_id	access_type
2013-05-15	2013-05-24		1	Took			F

KBart

NISO
How the information world CONNECTS

9

KBART Automation

2011...

The Future of KBART

Phase 1 – Universally accepted standardized publisher metadata, regularly distributed AND available on demand

Phase 2 – Broad adoption, Consortia, More content type coverage (eBooks, conference proceedings), Open Access

Phase 3? – Even more content types, automated delivery, institutional metadata????

KBART Phase III

Began March 2020

8 work items

KBART Standing Committee working in sub-groups on a few items at a time

Entire process, including initial draft, feedback, and final draft for publication = ~ 17 mos.

KBart

NISO
How the information world
CONNECTS

Introduce self.

Thanks for tuning in to our presentation during this time of COVID.

I hope you all find some comfort in a nice, simple presentation on planned updates to KBART to take your mind off the news.

Our timeline calls for an initial draft to be released 11 months into the process, which would be around February 2021.

Given the COVID-19 pandemic (including sickness on our Standing Committee), it is likely the timeline will be extended.

You should expect a draft sometime in 2021.

Work item 1: Clarify current recommendations

KBART Phase III will clarify the recommendations in the existing Recommended Practice.

- Expanded information on what file(s) to create and what metadata to include
- Clarifications and additional information on data fields
- More examples of correct implementation, preferably for every field or recommendation
- More guidance on handling:
    - Gaps in coverage for serials
    - Supplements for serials (which may have a different title but share an ISSN)
    - Title changes and title histories for serials (with respect to publisher ability to provide this data)
    - Handling of items withdrawn/no longer available for purchase

KBart

NISO
How the information world CONNECTS

13

Since the Phase II RP was published, the Standing Committee has fielded many questions asking about various recommendations.

With Phase III, we want to better communicate the intent of some of the recommendations and to generally make the KBART RP more readable and useful for the community.

What files to create: For example, whether to create separate ALL TITLES files for serials and monographs (instead of including them in same file), criteria on when to create a new file/package, etc.

More examples: For example, what to do when an issue or volume number is combined, as in 3/4; whether to require an end date when a journal is significantly behind in publication.

Title histories: For example, current recommendations for representing title histories (5.2, 6.6.2) might not work for content providers that don't have a unique identifier for preceding titles; maybe we need more than one recommendation depending on the situation.

Items withdrawn: What to do when some libraries have grandfathered access but other libraries have no access to titles in a collection.

Work on this item began in March 2020.

Work Item 2: Endorsement process

KBART Phase II endorsement process has only one tier for content providers.

For KBART Phase III, an overhaul of the endorsement process will be investigated:

- Varying levels of endorsement
- Endorsement of content providers vs. knowledge base vendors
- Branding and focus of program
- Role of KBART Registry in regard to how endorsement is communicated

Currently the endorsement process only has one tier for content providers. The lack of a multitiered approach causes several issues:
- Content providers who do achieve a "Gold Standard" should be rewarded, for which there is no provision in KBART Phase II.
- Content providers who are unable to attain 100% compliance due to technical limitations (such as inability to provide a journal's title history) may be dissuaded from attempting endorsement, even though the KBART Standing Committee may have been willing to issue an endorsement.
- It is not currently clear if knowledge base vendors can apply for endorsement or what standards should apply to them.*
- Adoption of KBART Automation is not considered currently as a criterion for endorsement.

After the release of Phase III, though we will encourage content providers to become Phase III-compliant, It will also be important to continue to acknowledge content providers who have achieved compliance with KBART Phase II, and we will need a way to indicate that.

We also want to look at the branding and the focus of the endorsement process as well as the role of the KBART Registry in regard to how endorsement is communicated.

[* Both the endorsement web page and the KBART registry indicate that the major KB vendors are endorsed.]

Work Item 3: Additional content types

KBART Phase II only provides holdings data for serials and monographs.

Content providers with multimedia and non-journal/non-book formats have no recommended way to communicate these holdings.

KBART Phase III will support additional content types:

- Textual (blogs, transcripts, websites, manuscripts, datasets, etc.)
- Non-textual (audio, video, images, etc.)

Since the last KBART RP revision in 2014, there has been an increase in popularity of non-textual content and textual content that doesn't fit the traditional serial and monograph formats that the KBART Recommended Practice assumes.

With no published recommendations to guide them, content providers who want to include multimedia and other non-book/non-journal formats in their KBART files often add extra fields to the end of the file* to identify such content, or they use the publication_type field (6.6.18), even though it only allows for the values "Serial" or "Monograph".

At best this causes confusion; at worst, content providers eliminate valuable data from their KBART title lists.

[* As per 6.6.1 "In order to supply additional information for other purposes not required by the KBART Recommended Practice, content providers are permitted to include any extra data fields *after* the last KBART utilized position (access_type in position 25).]

Work Item 3: Global content

Global content has little support in KBART Phase II.

KBART metadata does not identify translations of items or represent author names or titles in multiple languages.

KBART Phase III will offer improved support for global content.

The information supply chain has become more global, and information is increasingly shared on a worldwide basis.

A growing amount of scholarly content is produced in languages other than English and which do not use Latin characters, like CJK.

KBART Phase III will offer improved support for global content, taking into consideration:
- Content with non-Latin characters
- Translated titles
- Transliterated titles
- Expansion of author and editor name fields to include full names (because common surname might not be enough to uniquely identify item)
- Language of content itself

Work Item 4: File guide

Many content providers have an extensive catalog of content for sale (by content type, subject, geographic region, consortium, etc.) This results in a separate KBART file for each offering.

As content packages change, knowledge base vendors and librarians cannot easily keep track of what has been added, removed, or changed.

KBART Phase III will investigate requesting a document from content providers that serves as a guide to their files.

Such a guide, or file manifest, might include:
- Names of files delivered
- Collection name represented by the file
- Unique code for the collection
- Description of the collection
- Number of records in the collection
- Date file created or updated

It might also be possible to request a version history or the addition of add-delete-delta files to flag changes to the file since the last update. This would also be useful in supporting KBART Automation.

Work item 5: Sample license language

The Licensor will make available to third-party knowledge base providers an itemized holdings report that specifies the titles included in the Licensed Materials. The Licensor will use reasonable efforts to update itemized holdings reports as soon as is practicable when holdings information changes and will provide this information to knowledge base providers in a timely manner and to the Licensee on request. If the Licensed Materials include content covered by the NISO "Knowledge Bases and Related Tools (KBART) Recommended Practice", the Licensor will provide itemized holdings reports for the Licensed Materials in KBART-compliant format.

In addition, the Licensor will make available to third-party knowledge base vendors and Subscribing Institutions institution-specific holdings reports. If the Licensed Materials include content covered by the NISO "Knowledge Bases and Related Tools (KBART) Recommended Practice", the Licensor will make such holdings reports available for automated retrieving via an API that adheres to the requirements in the NISO "KBART Automation: Automated Retrieval of Customer Electronic Holdings" Recommended Practice.

KBART Phase II: Only mentions that the KBART Working Group has collaborated with consortia to provide guidance on statements within licenses and contracts relating to KBART compliance. [2.2.3]

A number of current model licenses in existence that mentioned KBART:
- California Digital Library model license (2016)
- LIBLICENSE model license (2015)
- Canadian Research Knowledge Network (CKRN) model license (2016)
- Jisc model license (2018)

Section 5.1 of KBART Automation (2019) has model license language, which we could possibly incorporate.

This is a DRAFT of model language that would support KBART files and KBART Automation.

Make it easier to include in licenses in order to encourage uptake of KBART.

Work item 6: Investigate alternate formats

Content providers have requested that KBART support file formats other than tab-delimited text files.

In KBART Phase III, we plan to identify the issues content providers are trying to resolve through alternate file formats.

Possible alternative file formats: XML, JSON

Would need to be *in addition to* tab-delimited text files

Key: Don't want to break anything by making changes.

Work item 6: Investigate alternate formats

Pros:

- Easier for knowledge base ingestion?
- Can handle multiple data elements (e.g. ISBNs, title histories)
- Better support for additional content types beyond journal and monograph
- XML, JSON can contain data about the file itself (e.g. collection name, date file created)
- Better support for APIs and KBART Automation

Cons:

- Current KB systems were developed to handle .txt files. Burden to develop support for other file types?
- Simplicity of .txt file and human-readability have been key to success of KBART thus far

KBart

NISO
How the information world
CONNECTS

20

Had productive discussion on this topic at NISO Plus in Baltimore in February 2020.

Work item 7: Article/chapter-level data

New business models:

- Some publishers are interested in selling content at an article or chapter level, e.g. on a specific topic.
- Hybrid Open Access journals contain some articles that are open and some that are only available to subscribers.

Current results in KBART

- Holdings data necessarily incorrect, resulting in false positives or negatives for user.
- KBART files that attempt to present very granular holdings are unwieldy.
- "Free" vs. "Paid" content communicated at journal title level only.

KBart

NISO
How the information world
CONNECTS

Raises questions:
- How can holdings information be properly communicated for part-of-whole items (articles, chapters, episodes, etc.)?
- Can this be done through KBART? Should it be?
- Is access at this level a role for knowledge bases? Is it a role for discovery systems?
- What possible solutions exist, and who needs to be involved in the discussion?

Work item 7: Article/chapter-level data

KBART Phase III will create a roadmap for communication of article/chapter-level holdings data.

- What work needs to be done in the future to realize support for article/chapter-level data?
- What groups and technical experts may be needed to accomplish this task?

Support for article and chapter-level data is *not* solvable within KBART Phase III, but we want to lay the foundation for the consideration of this problem by multiple groups.

We will likely need a data format that works in conjunction with KBART, but which will require its own process.

There are a number of NISO and other affiliated/related groups that work on areas that are adjacent to, overlap, inform, or are dependent on KBART. We hope to collaborate with them to address this issue.

Work item 8: KBART mission

KBART was originally created to support accuracy in OpenURL linking.

KBART Phase II expanded KBART to address consortia holdings, open access content, and e-books and conference proceedings.

Now, KBART is being used in unintended ways not imagined when KBART was first drafted.

This is largely due to the success of KBART: its simplicity and wide adoption in the information supply chain.

One thing to note is that we're very clear that KBART is for communicating **holdings** data, not descriptive metadata. For example, some providers include subjects in their KBART files, in an extra column, and that's fine, but we would not consider adding something like a subject field to the RP.

While we want to expand the KBART mission* to reflect KBART's many uses, we also want to guard against mission-creep.

Work on this item began in March 2020.

[* No actual "mission" section in current RP; just section on "Original Charge and Scope".]

KBART Phase III: How can you get involved?

Take our short survey on: "What are the top three priorities that you are hoping that a revised KBART Recommended Practice would address?" at https://bit.ly/KBART2019

Or, contact us by email: kbart@niso.org

Join the interest group mailing list: https://groups.niso.org/lists/kbart_interest/

Respond to the KBART Phase III draft during the public feedback period

We hope that by updating KBART to clarify current recommendations and expand it to include more content types, global content, and methods of data transmission, KBART will be more useful and relevant across the electronic content that libraries license and access in 2020 and beyond.

This, in turn, will incentivize content providers to adopt the KBART RP, and that in turn will improve access by end users to these valuable information resources.

Thank you.

Any questions for us?