

ALA American Library Association

# THE KNOWLEDGE BASE AT THE CENTER OF THE UNIVERSE

Kristen Wilson

**Library Technology Reports**

Expert Guides to Library Systems and Services

AUG/SEPT 2016

Vol. 52 / No. 6

ISSN 0024-2586

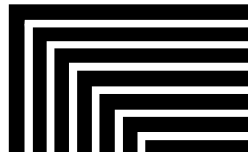
# Library Technology

R E P O R T S

Expert Guides to Library Systems and Services

## **The Knowledge Base at the Center of the Universe**

*Kristen Wilson*



**ALA TechSource**  
[alatechsource.org](http://alatechsource.org)

American Library Association

# Library Technology R E P O R T S

ALA TechSource purchases fund advocacy, awareness, and accreditation programs for library professionals worldwide.

**Volume 52, Number 6**

**The Knowledge Base at the Center of the Universe**

ISBN: 978-0-8389-5975-6

**American Library Association**

50 East Huron St.  
Chicago, IL 60611-2795 USA  
alatechsource.org  
800-545-2433, ext. 4299  
312-944-6780  
312-280-5275 (fax)

**Advertising Representative**

Patrick Hogan  
phogan@ala.org  
312-280-3240

**Editor**

Patrick Hogan  
phogan@ala.org  
312-280-3240

**Copy Editor**

Judith Lauber

**Production**

Tim Clifford and Alison Elms

**Cover Design**

Alejandra Diaz

*Library Technology Reports* (ISSN 0024-2586) is published eight times a year (January, March, April, June, July, September, October, and December) by American Library Association, 50 E. Huron St., Chicago, IL 60611. It is managed by ALA TechSource, a unit of the publishing department of ALA. Periodical postage paid at Chicago, Illinois, and at additional mailing offices. POSTMASTER: Send address changes to *Library Technology Reports*, 50 E. Huron St., Chicago, IL 60611.

Trademarked names appear in the text of this journal. Rather than identify or insert a trademark symbol at the appearance of each name, the authors and the American Library Association state that the names are used for editorial purposes exclusively, to the ultimate benefit of the owners of the trademarks. There is absolutely no intention of infringement on the rights of the trademark owners.



Copyright © 2016  
Kristen Wilson  
All Rights Reserved.

## About the Author

**Kristen Wilson** is Associate Head of Acquisitions and Discovery at the North Carolina State University Libraries. She manages the department's serials unit and has been involved in the development of a local electronic resources management system and the Global Open Knowledgebase (GOKb). She has published on the topics of knowledge bases, electronic resources management, and workflow analysis in journals such as *Serials Review* and *Library Resources and Technical Services*. She holds a BA in English and Journalism from Lehigh University and an MSLIS from Syracuse University.

## Abstract

More than fifteen years after its initial appearance, the electronic resources knowledge base has come into its own as a tool that touches nearly every area of library management. And the knowledge base continues to evolve, expanding into areas such as APIs, open data, community contribution models, and integration with next-generation systems. This *Library Technology Report* will analyze the impact of knowledge bases on library management practices and explore new directions and trends for these tools. The report will trace the evolution of the knowledge base, provide context for knowledge base creation and maintenance, and explore areas of innovation including use in library services platforms, integration with external tools and services, and collaborative knowledge base projects.

## Get Your *Library Technology Reports* Online!

Subscribers to ALA TechSource's *Library Technology Reports* can read digital versions, in PDF and HTML formats, at <http://journals.ala.org/ltr>. Subscribers also have access to an archive of past issues. After an embargo period of twelve months, *Library Technology Reports* are available open access. The archive goes back to 2001.

## Subscriptions

[alatechsource.org/subscribe](http://alatechsource.org/subscribe)

# Contents

<b>Chapter 1—Knowledge Base Evolution</b>	<b>5</b>
The Origin of Knowledge Bases	5
Beyond OpenURL	6
Knowledge Base Structure	7
Conclusion	8
Notes	8
<b>Chapter 2—Building a Knowledge Base</b>	<b>9</b>
The Supply Chain	9
Building a Knowledge Base	9
When It Goes Wrong	10
Knowledge Bases and Related Tools (KBART)	11
Conclusion	11
Notes	12
<b>Chapter 3—Knowledge Bases and Library Services Platforms</b>	<b>13</b>
EBSCO's EBSCONet, Usage Consolidation, and ERM Essentials	13
OCLC's WorldShare Management Services	14
ProQuest's Intota and Alma	15
Other Services	16
Conclusion	16
Notes	16
<b>Chapter 4—New Uses for Knowledge Bases</b>	<b>18</b>
Enhanced Knowledge Base Data	18
Knowledge Base APIs	19
Integrations with Other Products	20
Delivery of Library-Specific Holdings	20
Decision Support	21
Conclusion	22
Notes	22
<b>Chapter 5—Collaborative Knowledge Bases</b>	<b>23</b>
Community-Managed Knowledge Bases	23
National Knowledge Bases	25
Conclusion	26
Notes	27
<b>Chapter 6—Knowledge Base Profiles</b>	<b>28</b>
Commercial Knowledge Bases	28
Open Knowledge Bases	32



# Knowledge Base Evolution

The electronic resources knowledge base began humbly, an unglamorous piece of infrastructure often overlooked in the excitement surrounding high-profile discovery services. But more than fifteen years after its initial appearance, the knowledge base has come into its own as a tool that touches nearly every area of the library management sphere. And the knowledge base continues to evolve, expanding into areas such as APIs, open data, community contribution models, and integration with next-generation library services platforms (LSP). This issue of *Library Technology Reports* will analyze the impact of knowledge bases on library management practices and explore new directions and trends for these tools.

Chapter 1 provides a basic introduction to knowledge base terminology and functionality and draws on the published literature to describe the product's evolution. Chapter 2 examines the process of creating and maintaining a knowledge base and the role of key players across the supply chain.

Chapters 3, 4, and 5 will focus on areas of innovation for knowledge bases. Chapter 3 describes the use of knowledge bases within the emerging class of management tools known as library services platforms. In chapter 4, extensive interviews with vendors, content providers, and librarians inform a discussion of new directions in knowledge base development and use. Chapter 5 explores the trend toward encouraging greater collaboration and openness through open-source, community, and national knowledge base projects.

Chapter 6 provides an overview of the current product landscape. A listing of the major commercial and open-source knowledge bases is accompanied by short descriptions of each product provided by the company or organization that maintains it.

## The Origin of Knowledge Bases

The history of the knowledge base is closely entwined with the development of the OpenURL link resolver in the late 1990s. The OpenURL resolver made its conceptual debut in a series of articles published in 1999 by Van de Sompel and Hochstenbach. The authors addressed the appropriate copy problem by describing an approach to dynamic linking.<sup>1</sup> Rather than attempt to hard-code links from a source citation to specific copies of an article, they developed a prototype tool that created links to an appropriate copy on the fly, using information provided by two sources: the citation being viewed and a store of information about content providers and how to link to their resources.<sup>2</sup> The tool, which was called SFX, was acquired by Ex Libris in 2000 and soon after released as the first commercial link resolver. Early descriptions of SFX hinted at the concept that would eventually evolve into today's knowledge base. In an article explaining emerging OpenURL technology to a general audience, Walker simply mentioned that SFX includes a database that describes an institution's collection and the types of services it chooses to provide to its users.<sup>3</sup>

At the same time that OpenURL development was bringing about one early version of the knowledge base, the same concept was evolving as part of another tool. In 2000, a new company called Serials Solutions began offering a service that tracked the content of aggregator packages and generated a localized A-to-Z list of titles based on a library's subscriptions.<sup>4</sup> The underlying metadata surrounding the Serials Solutions service—information describing an institution's collection and how to access it—ended up being very similar to that needed to power an OpenURL resolver.

The synergy was so great, in fact, that within a few years Serials Solutions began to offer its own link resolver and SFX its own A-to-Z list. Today, these two companies have become one, following ProQuest's acquisition of first Serials Solutions in 2011 and then Ex Libris in 2016. In the decades following the initial development of these products, many more companies across the library ecosystem began to offer their own competing solutions.

Wider adoption of tools relying on knowledge bases also brought about greater scrutiny of the quality of data provided and the effort needed to maintain a library's local holdings. Early writings on knowledge bases vary widely in their assumptions about how easy or difficult this process would turn out to be. Caplan and Arms were impressively apt in their assessment of the problems of scale that would plague knowledge base maintenance from both the vendor and library perspectives. They still missed the mark, however, in their assumption that these difficulties would prevent the successful implementation of global knowledge bases as a component of link resolver products.<sup>5</sup> Walker's opposite assessment that "it is clear that these tasks have relatively insignificant resource implications" seems comically naïve in the current environment.<sup>6</sup>

By 2006, the true implications of a reliance on knowledge bases began to crystalize. Wakimoto, Walker, and Dabbour identified the accuracy and completeness of the knowledge base as a key determinant of the quality of a link resolver. They also noted the extent to which librarians have begun to contribute their expertise back to the link resolver vendors, citing one librarian who reported roughly thirty errors to Ex Libris each month.<sup>7</sup> In her issue of *Library Technology Reports* the same year, Grogg urged readers to consider knowledge base quality as a top factor in the decision about which knowledge base to purchase.<sup>8</sup> At this point, the knowledge base had become established as core library infrastructure requiring both time and effort to manage and underpinning many of a library's most visible services.

While it's impossible to definitely state the number or percentage of libraries currently using knowledge base-driven products, the numbers that are available suggest very widespread adoption. In response to the profile questionnaire for this report, three of the largest library systems providers—EBSCO, OCLC, and ProQuest—reported a combined 11,700 libraries using products that rely on their knowledge bases. Ex Libris's corporate website lists another 5,600 total customers, many of which are likely relying on its knowledge base.<sup>9</sup> Several smaller vendors offer knowledge base-powered products as well, and many open-source knowledge bases are used on an informal and thus unmeasurable basis.

## Beyond OpenURL

While knowledge bases may have evolved to support specific tools like OpenURL link resolvers, the wide-ranging usefulness of their data has made them prime infrastructure on which to build new services. In the years since their initial development, knowledge bases have come to integrate with a new wave of library tools, including electronic resources management systems (ERMSs), discovery products, and library services platforms (LSPs).

The ERMS was the earliest of the second wave of tools to take advantage of knowledge base data. These systems aim to provide a suite of services specifically scoped toward managing electronic journals and books—services that are significantly not part of the traditional integrated library system (ILS), which was designed with a print world in mind. Typical features of an ERMS include management of license agreements, contact information, administrative metadata for e-resources platforms, and usage statistics. Underlying all of these functions is the ability for a library to track its collection and create linkages between a resource and the ERMS components that relate to it. The knowledge base is a logical source of this metadata, as it already contains structured data about a library's holdings and in many cases is already being maintained by the library to support discovery tools.

The ERMS is now largely considered to be a stop-gap on the road to the development of the LSP, which attempts to unite the functions of the knowledge base, ERMS, and ILS under one umbrella. Breeding clarified that LSPs do not necessarily contain a consistent set of functionality across different vendors' products, but rather are defined by a unified approach to managing all resource types and providing flexible services such as APIs that allow for interoperability and custom development.<sup>10</sup> The role of the knowledge base within the library services platform is still evolving as these solutions gain a foothold in the market. Chapter 3 of this report will address new developments in this area more specifically.

Patron-facing discovery products, in the form of unified search indexes, have also benefited from the use of a knowledge base. The knowledge base plays a key role in these discovery products in two ways. First, it allows libraries to scope the huge sets of search results returned by discovery tools to only items in their own collections. Second, it continues in its traditional role supporting a link resolver. While discovery services index the full text of articles and book chapters, their agreements with publishers prevent them from actually exposing the full text. So users must still rely on reference linking to get from their source citations to the content itself. Much of this has been done through traditional OpenURL resolution,

On this date	This/These titles	Changed to This/These titles	
1975-01-01	• Physica	• Physica A: Statistical Mechanics and its Applications • Physica B+C	Delete
1988-01-01	• Physica B+C	• Physica B: Condensed Matter • Physica C: Superconductivity	Delete

**Figure 1.1**  
Title level metadata in the Global Open Knowledgebase (GOKb) includes detailed publication information.

although that practice is rapidly giving way to new direct-linking technology, which leverages the metadata in the unified index to create links, rather than constructing them based on information in the source citation. Ironically, the very technology that helped launch the knowledge base may be eroding, while the knowledge base itself lives on in other contexts.

## Knowledge Base Structure

In response to the needs of the tools described above, knowledge bases have evolved a fairly consistent structure and data model. It's worth briefly addressing the general model in a bit more detail, as well as the tools that allow librarians to interact with the knowledge base in an administrative capacity.

Unlike traditional bibliographic records, which aim to describe publications at a work level, knowledge bases focus on describing holdings—the specific version of a work that a library can purchase and provide access to. This approach is what makes knowledge base data so useful: it can help a library describe and manage its collections in a practical way that models the reality of how resources are sold and accessed. Knowledge bases collect and track the entities that together define the holding. The work-level title is of course still an essential piece of this concept. Knowledge bases store a lot of important metadata related to titles, including variant and abbreviated titles; ISSNs, IBSNs, and other unique identifiers; publisher names; and where appropriate, additional data like subject headings, LC classes, authors, title histories, and editions (see Figure 1.1).

The titles in a knowledge base are grouped into packages that describe the way resources are purchased (see Figure 1.2). Packages might represent bundles of content sold by the publisher such as subject collections, back files, and big deals. Aggregator packages describe collections of content packaged and sold as databases by third parties like EBSCO and ProQuest. And many packages simply describe master lists—all of the titles provided by a publisher or content provider. In the case of the smallest publishers, a master list package may contain only a single title.

Title	Content Availability	Identifiers	In My Holdings
<a href="#">Academic Pediatrics</a>	Full Text: From 2009-01 to present Full Text: Volume 9, Issue 1	OCN: 782077887 ISSN: 1676-2859 <a href="#">Website</a>	🟢 Held by my library
<a href="#">Academic Radiology</a>	Full Text: From 2007-01 to present Full Text: Volume 14, Issue 1	OCN: 49419824 ISSN: 1076-6332 <a href="#">Website</a>	🟢 Held by my library
<a href="#">Addictive Behaviors</a>	Full Text: From 2007-01 to present Full Text: Volume 32, Issue 1	OCN: 137345754 ISSN: 0306-4603 <a href="#">Website</a>	🟢 Held by my library

**Figure 1.2**  
A package record in the OCLC World Cat Knowledge Base displays a list of titles and holdings and allows users to search and filter the contents.

Title Name	Search	Show:	Selected and Not Selected	Limit To:
22 Titles Found				
<b>Tetrahedron</b>				
ISSN (Print) : 0040-4020    ISSN (Online) : 1464-5416				
Publisher: Pergamon    Resource Type: Journal    Subject: Organic Chemistry				
Package Name	Coverage	Vendor		
<a href="#">Access to Research for Development and Innovation (ARDI)</a>	01/01/1995 - Present	World Intellectual Property Organization		
<a href="#">AGORA</a>		FAO (Food and Agriculture Organization)		
<a href="#">CIBER-CASPUR</a>	01/01/1957 - Present	CASPUR		
<a href="#">Clea ScienceDirect</a>	01/01/1995 - Present	Elsevier Science		
<a href="#">Elsevier - WKC</a>	01/01/1995 - Present	Elsevier Science		

**Figure 1.3**  
The EBSCO knowledge base displays a list of holdings that represent the various ways a title can be purchased.

In most knowledge bases, the combination of a title and a package makes up a holding. The holding record contains metadata that aids in access and management of a purchase—the years of coverage provided with the purchase, the URL where the resource can be accessed, and in some cases management information like whether or not the content is open-access (see Figure 1.3). In traditional knowledge bases, holdings can be activated; essentially they are given a tag that states “my library owns this title, as part of this package, with this coverage range and URL.” That information can be used by related systems to help end users access resources and librarians manage their collections.

Knowledge bases can also contain a range of other components that relate to the resources being described, including organizations, providers, and platforms. These record types store additional metadata about the entities involved in making e-resources available and also help collocate resources based on a common provider or platform. Because there is no industry standard data model for knowledge bases, the use of these entities varies between products.

Knowledge bases push their data out to many other systems, but they almost always offer a separate administrative interface that allows librarians to interact with the data and configure system settings (see Figure 1.4). They can search for known items and browse by exploiting links between various entities. For knowledge bases that can be localized to represent an institution's holdings, special fields allow titles to be included



<b>Title</b>	Comparative literature
<b>Provider</b>	Duke University Press
<b>Database</b>	Duke University Press
<b>Type</b>	Journal
<b>ISSN</b>	0010-4124
<b>Publisher</b>	University of Oregon
<b>Status</b>	Subscribed ▼
<b>Display In:</b>	<input checked="" type="checkbox"/> 360 Core
	<input checked="" type="checkbox"/> 360 Link
	<input checked="" type="checkbox"/> 360 MARC Updates
	<input checked="" type="checkbox"/> Summon
<b>Default Coverage Dates</b>	01/01/2000 -
<b>Custom Coverage Date From</b>	<input type="text"/>
<b>Custom Coverage Date To</b>	<input type="text"/>
<b>Output Coverage Date(s)</b>	01/01/2000 - Present
<b>Default URL</b>	<a href="http://complit.dukejournals.org">http://complit.dukejournals.org</a>
<b>Custom URL</b>	<input type="text"/>
<b>Output URL</b>	<a href="http://proxying.lib.ncsu.edu/index.php?url=http://complit.dukejournals.org">http://proxying.lib.ncsu.edu/index.php?url=http://complit.dukejournals.org</a>

**Figure 1.4**

Administrative functions in the ProQuest Knowledgebase allow users to set the status of a holding, customize coverage dates and URLs, and control display options in public facing tools.

or excluded from specific services. These knowledge base interfaces are aimed at administrative users and are never seen by library patrons.

## Conclusion

While knowledge bases were initially created as a byproduct of OpenURL link resolvers and A-to-Z lists, they have evolved into useful tools in their own right. In their modern context, knowledge bases provide libraries with an inventory of electronic book and journal holdings and describe the materials that a library has purchased at a more granular level than the traditional bibliographic record. Knowledge base data supports a wide variety of discovery tools, from the original link resolvers to new unified search platforms. Knowledge bases are also used to support management needs throughout the e-resources life cycle in

areas such as licensing, usage statistics, and resource sharing. It's safe to say that the knowledge base has truly become the center of the management universe for academic and research libraries.

## Notes

1. Herbert Von de Sompel and Patrick Hochstenbach, "Reference Linking in a Hybrid Library Environment, Part 1: Frameworks for Linking," *D-Lib Magazine* 5, no. 4 (April 1999), [www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt1.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt1.html).
2. Herbert Von de Sompel and Patrick Hochstenbach, "Reference Linking in a Hybrid Library Environment, Part 1: SFX, a Generic Linking Solution," *D-Lib Magazine* 5, no. 4 (April 1999), [www.dlib.org/dlib/april99/van\\_de\\_sompel/04van\\_de\\_sompel-pt2.html](http://www.dlib.org/dlib/april99/van_de_sompel/04van_de_sompel-pt2.html).
3. Jenny Walker, "Open Linking for Libraries: The New OpenURL Framework," *New Library World* 102, no. 4/5 (2001): 127-34, <http://dx.doi.org/10.1108/03074800110390482>.
4. Paula D. Watson, "E-Journals: Access and Management," *Library Technology Reports* 39, no. 2 (March 2003): 44-68.
5. Priscilla Caplan and William Y. Arms, "Reference Linking for Journal Articles," *D-Lib Magazine* 5, no. 7/8 (July/August 1999), [www.dlib.org/dlib/july99/caplan/07caplan.html](http://www.dlib.org/dlib/july99/caplan/07caplan.html).
6. Walker, "Open Linking for Libraries," 132.
7. Jina Choi Wakimoto, David S. Walker, and Katherine S. Dabbour, "The Myths and Realities of SFX in Academic Libraries," *Journal of Academic Librarianship* 32, no. 2 (March 2006): 127-36, <http://dx.doi.org/10.1016/j.acalib.2005.12.008>.
8. Jill E. Grogg, "Linking and the OpenURL," *Library Technology Reports* 42, no. 1 (January-February 2006): 24, <http://dx.doi.org/10.5860/ltr.42n1>.
9. "Our Vision," Ex Libris, 2015, [www.exlibrisgroup.com/category/Our\\_Vision](http://www.exlibrisgroup.com/category/Our_Vision).
10. Marshall Breeding, "Library Services Platforms: A Maturing Genre of Products," *Library Technology Reports* 51, no. 4 (May/June 2015), <https://journals.ala.org/ltr/issue/view/509>.

# Building a Knowledge Base

The specifics of the process that knowledge base suppliers use to collect, normalize, and maintain their data may be of interest to a broader audience for two reasons. First, librarians interact with these tools intimately and invest a lot of time and effort aiding vendors in keeping knowledge bases up to date. A better understanding of how this work is done provides context for these activities. Second, as knowledge bases begin to reach beyond vended products into national, consortial, and institutional arenas, librarians may soon find themselves more deeply embedded in the process of knowledge base work.

## The Supply Chain

Knowledge base metadata originates with the provider of an e-resource—usually the original publisher or a third-party content aggregator. These organizations create data files describing the products they sell. Each file generally represents a salable package made up of a specific set of titles. The lists may include e-journals, e-books, or a mix of both. For each title, the file provides the information needed to identify and access its content, such as unique identifiers, coverage dates, and URLs. This data, which is most often made available as a simple Excel or tab-delimited file, may be made publicly available on a provider's website or provided to knowledge base suppliers through FTP sites, e-mail, or other methods.

Knowledge base suppliers collect title list metadata from content providers and load it into their products, performing quality checks and normalization processes along the way to improve accuracy. These suppliers then distribute the data to libraries through their knowledge base software. Libraries,

meanwhile, collect data about their local purchases from vendors, publishers, and subscription agents. They use their own records to identify the packages and titles they have purchased in their local implementation of a knowledge base. Once the appropriate resources are activated, the accompanying metadata is pushed out for use across the library's systems and services. Librarians often attempt to close the loop in this process by reporting changes and corrections to the data back to their knowledge base vendor or the content provider itself.<sup>1</sup>

## Building a Knowledge Base

From the perspective of the group building a knowledge base, the first major step is collecting data from content providers. For commercial knowledge bases attempting a comprehensive list of scholarly publications, this process can be an enormous undertaking. Jackie Fahmy, knowledge base product analyst at OCLC, reported in an interview with the author that the WorldCat Knowledge Base contains data from more than 5,000 providers.<sup>2</sup> Oliver Pesch, chief product strategist at EBSCO, cites the numbers for EBSCO's Global Knowledge Base at more than 1,400 providers totaling more than 10,000 unique collections.<sup>3</sup> Smaller knowledge bases are often unable to achieve such a large scale and must limit their scope accordingly. Niche products like CUFTS, run by Simon Fraser University, and GoldRush, based at the Colorado Alliance of Research Libraries, scope their coverage to those collections specifically needed by their customers. National projects may address only publications native to their countries or packages purchased nationally at a consortial level. The Global Open

Knowledgebase (GOKb), an open-source, community-managed project, has begun work by focusing on priority packages and master lists, with a hope of increasing scale through evolving technology and partnerships.

While some content providers make suitable title lists freely available on their websites, others require special arrangements for knowledge base suppliers to access the data. Data can be delivered through several different mechanisms including websites, FTP sites, APIs, and occasionally even e-mail. FTP sites and APIs are ideal, as they allow providers to build harvesters that automatically pull in files, rather than forcing a human to visit the site and download the file manually.

Once a data pipeline is established, the next stage is normalizing and cleaning the data before ingesting it into the production version of the knowledge base. These validation processes aim at achieving consistency across all the resources represented in the knowledge base. Examples of the types of work done at this stage include checking for required fields, properly formatting data like dates and identifiers, and analyzing changes to the files from one update to the next. Additionally, knowledge base suppliers must compare the incoming data against the existing knowledge base. When discrepancies arise, they decide which version of the information is correct and choose to either replace the existing data or disregard the incoming change.

Much of the normalization work is automated, but a human element is still a key part of the process. OCLC, for instance, sets a 5 percent threshold for data changes in a content provider's file. If the threshold is exceeded, the file is flagged for manual review.<sup>4</sup> EBSCO monitors nearly all types of changes to the incoming data, as well as flagging new and dropped titles for review. Once all the review events are addressed, the file can be fully ingested for use in EBSCO's products.<sup>5</sup> The GOKb project divides data exceptions into groups of pre- and post-ingest tasks that are reviewed by its community contributors. Pre-ingest tasks focus on normalizing data before it becomes part of the knowledge base. The post-ingest tasks address discrepancies between old and new data, as well as anomalies that don't necessarily prevent ingest, but could cause problems for users down the road.

Because electronic resources products evolve so rapidly, the data collection and ingest process must be repeated on a regular and frequent basis. The Knowledge Bases and Related Tools (KBART) code of practice recommends that providers issue monthly updates, and many of the larger providers seem to be adhering to this schedule, or even exceeding it.<sup>6</sup> Smaller providers may update their files less often, but they may also have fewer changes to their metadata to warrant a higher frequency. As a result, knowledge

base suppliers must be aware of the general schedule used by each provider so that they can continuously harvest, process, and load files to keep their products up-to-date.

To this end, commercial knowledge base providers employ sizable staffs whose job it is to maintain the knowledge base. Breeding reported in 2012 that the four major knowledge base suppliers (Ex Libris, Serials Solutions, EBSCO, and OCLC) employed between eight and twenty-nine full-time employees involved in knowledge base maintenance.<sup>7</sup> A smaller supplier, like CUFTS, has about five staff members who regularly work on the knowledge base, though none of them are full time. GoldRush relies on library school students working part-time to handle its file processing. GOKb employs one full-time staff member and relies on volunteer effort from project partners to help review errors and participate in data enhancement activities.

## When It Goes Wrong

The validation stage of knowledge base creation is essential because the data being consumed is prone to errors—due to its complexity and its status as a secondary artifact of the publishing process. “The data we get isn't always clean, pristine data,” said Yvette Diven, product manager lead for management solutions at ProQuest. “This metadata can be a byproduct of something that a provider does. If they're focused on publishing e-journals or e-books, this metadata can be a byproduct rather than the main product.”<sup>8</sup>

The types of errors commonly found in knowledge bases are well documented. In an early analysis, Chen noted that content providers often failed to update their metadata frequently enough to capture titles added and dropped from their collections. She also provided several examples of data errors at the title level, such as incorrect coverage dates and URLs.<sup>9</sup> Cullen described similar issues broken down into a useful list that includes missing titles, titles listed in error, wrong identifiers for titles (ISSN, ISBN), incorrect coverage information, and incorrect embargo information.<sup>10</sup> Another error type frequently seen in knowledge bases involves the correct representation of serial titles over their life span, including title changes and transfers between publishers. While the introduction of the KBART code of practice has helped promote more frequent updates, metadata problems continue to be an issue for knowledge base suppliers and users, as the entire supply chain struggles to keep up with the volume of changes.

The consequence of bad knowledge base data can be felt across the internal and external operations of the libraries relying on it. The sharpest pain point is for end users of link resolvers and discovery tools, who may be incorrectly told their library has

no access to the article they're searching for—or, perhaps worse, directed to a resource they believe should be available, only to be faced with a pay wall or error message. Librarians also feel the frustration caused by knowledge base errors, which can make it difficult for them to manage their collections, reconcile title lists, analyze usage, and troubleshoot end user problems.

Because these errors have the strongest impact on librarians and library patrons, customers also play a role in helping to maintain the quality of knowledge bases. Every knowledge base supplier I spoke with provides a way for customers to report errors discovered through real-world use of the data. The suppliers then review these error reports, confirm proposed changes with the content provider, and edit the knowledge base if appropriate. While user participation in knowledge base maintenance certainly benefits users themselves and their knowledge base suppliers, Cullen rightly pointed out that the current model can also lead to inefficiencies. Librarians from different institutions will identify and report the same errors to various suppliers. And at times, suppliers are less likely to prioritize user error reports, leading to delays in these changes being applied.<sup>11</sup>

What's needed is additional effort to close the gaps in the supply chain by fixing problems at their source and building environments for greater collaboration. Most knowledge base providers already address the first aspect of this need by communicating known errors back to content providers whenever possible. Chapter 5 of this report will examine national and community efforts to improve the supply chain at a more grassroots level.

## Knowledge Bases and Related Tools (KBART)

The biggest challenges surrounding knowledge base maintenance include the sheer volume of data that must be processed, the need to provide timely information, and the task of modeling complex and ever-changing collections of resources. The KBART code of practice was created to address these challenges by defining effective participation in the supply chain. The foundation for KBART was originally proposed by Culling in his 2007 report to UKSG, which identified the need to establish transparent guidelines for how best to format, deliver, and consume knowledge base data.<sup>12</sup> The original KBART working group was formed as a joint venture between UKSG and NISO in 2007. In 2010, the original recommended practice was released, followed by a Phase II revision by NISO in 2014.<sup>13</sup>

While the initial exploration for KBART covered broad ranging topics—including OpenURL syntax and compliance, the role of subscription agents, and

the handling of e-journal title changes—the code of practice that emerged has so far focused mainly on the supply of title list files from content providers to knowledge base suppliers. KBART defines the method, frequency, and format of data exchange, along with a set of twenty-five fields to be included in each file. A KBART-compliant title list is a simple tab-delimited file, it can be delivered via a dedicated web page or FTP site, and the fields are all quite straightforward and eye readable.

The KBART standing committee continues work on the initiative, focusing on education and outreach. The committee conducts training workshops for implementers of KBART, provides endorsement for organizations that have demonstrated successful adoption, and maintains a registry of KBART-compliant file sources and contacts.<sup>14</sup> The KBART website currently lists forty-six endorsed organizations, and many unendorsed content providers use the code of practice informally.

The success of KBART has led to some discussion of additional uses and improvements for the code of practice. In a 2014 article, EBSCO's Oliver Pesch identified several new use cases for KBART, including the exchange of KBART data between vendors to allow customers to mix and match products; identification of lendable items for document delivery; and delivery of custom KBART files describing an individual library's holdings.<sup>15</sup> New uses for KBART and knowledge base data in general will be discussed more fully in chapter 4.

## Conclusion

The knowledge base supply chain is really a complex web of players who create, consume, enhance, and make use of title list metadata. The process of collecting this information and transforming it into an accurate, consistent knowledge base is a monumental undertaking that can be accomplished at scale by only the largest vendors. At the same time, smaller players, including national, consortial, and open-source knowledge bases, focus on niche areas appropriate to their user bases—and in the process become experts on certain types of content. Individual libraries retain a key role in the supply chain by correcting and improving data issues discovered through real-world use. Together, these groups have managed to put together a system for the creation and maintenance of knowledge bases that has been quite successful—especially when judged against some of the early doubts about the products' feasibility.

Still, areas of inefficiency persist. Each of the large knowledge base providers essentially duplicates the efforts of the others. They all collect the same data and must handle the same errors and inconsistencies.

Closing the loop with content providers also remains a challenge. While libraries and knowledge base suppliers make some efforts to improve data at its source, publishers often lack the resources to acknowledge these changes or implement them in a meaningful way. And while KBART and other standards have made a big impact on the efficiency of data delivery, other areas of the supply chain—such as use of ISSNs and handling of title changes and transfers—could still benefit from additional codification. The following chapters of this report illustrate the extent to which these challenges are recognized across the supply chain and describe many new initiatives that aim to meet them.

## Notes

1. James Culling, *Link Resolvers and the Serials Supply Chain*, final project report for UKSG (Oxford, UK: Scholarly Information Strategies, May 21, 2007), [www.uksg.org/projects/linkfinal](http://www.uksg.org/projects/linkfinal).
2. Jackie Fahmy (knowledge base product analyst at OCLC) in discussion with the author, October 2015.
3. Oliver Pesch (chief strategist at EBSCO) in discussion with the author, November 2015.
4. Fahmy discussion.
5. Pesch discussion.
6. KBART Phase II Working Group, *Knowledge Bases and Related Tools (KBART) Recommended Practice*, NISO Recommended Practice RP-9-2014 (Baltimore, MD: National Information Standards Organization, 2014), [www.niso.org/apps/group\\_public/download.php/12720/rp-9-2014\\_KBART.pdf](http://www.niso.org/apps/group_public/download.php/12720/rp-9-2014_KBART.pdf).
7. Marshall Breeding, “E-Resource Knowledge Bases and Link Resolvers: An Assessment of the Current Products and Emerging Trends,” *Insights* 25, no. 2 (July 2012): 173–82.
8. Yvette Diven (product manager lead for management solutions at ProQuest) in discussion with the author, October 2015.
9. Xiaotian Chen, “Assessment of Full-Text Sources Used by Serials Management Systems, OpenURL Link Resolvers, and Imported E-Journal MARC Records,” *Online Information Review* 28, no. 6 (2004): 428–34, <http://dx.doi.org/10.1108/14684520410570553>.
10. Culling, *Link Resolvers and the Serials Supply Chain*, 28.
11. *Ibid.*, 31.
12. Culling, *Link Resolvers and the Serials Supply Chain*.
13. KBART Phase II Working Group, *Knowledge Bases and Related Tools (KBART) Recommended Practice*, NISO Recommended Practice RP-9-2014 (Baltimore, MD: National Information Standards Organization, 2014), [www.niso.org/apps/group\\_public/download.php/12720/rp-9-2014\\_KBART.pdf](http://www.niso.org/apps/group_public/download.php/12720/rp-9-2014_KBART.pdf).
14. KBART Registry, NISO, accessed April 29, 2016, <https://sites.google.com/site/kbartregistry>.
15. Oliver Pesch, “The KBART’s Potential beyond OpenURL Linking,” *Serials Librarian* 67, no. 3 (2014): 231–39, <http://dx.doi.org/10.1080/0361526X.2014.960643>.

# Knowledge Bases and Library Services Platforms

One of the most exciting new uses of the knowledge base is its central place in the current generation of library services platforms (LSPs). These systems aim to integrate functionality traditionally distributed across siloes, specifically the discovery knowledge base, the integrated library system (ILS), and the electronic resources management system (ERMS). Oliver Pesch, chief product strategist at EBSCO, sums up the way that bringing these environments together around a central knowledge base benefits libraries:

For things to function properly—acquisitions and circulation and discovery and linking and usage and all of those pieces—you need to be consistent in how you talk about the resources that are in your collection. And the best way of doing that is to have some fairly rich central knowledge base in the middle that, if nothing else, is acting as an identity broker. You can take and enhance that thing that you decided to purchase. You're not copying that thing each time in a new system, but supplementing it. We see the knowledge base as pretty important in that world.<sup>1</sup>

By using the knowledge base to create consistent, reusable representations for electronic resources holdings, these new systems promise to offer libraries new levels of efficiency, interoperability, and automation. And while there is still work to be done before any one vendor fully realizes the LSP vision, several of the larger players have already made significant progress. These systems offer a taste of the impact that a knowledge base-centered system will have on the library management space.

## EBSCO's EBSCONet, Usage Consolidation, and ERM Essentials

While EBSCO doesn't offer a single, unified LSP, its suite of tools demonstrates an understanding of how to use a central knowledge base to achieve consistency and interoperability across related products. EBSCO's global knowledge base supports its discovery products, including EBSCO Discovery Service and the LinkSource OpenURL resolver. As a subscription agent and content provider, the company has begun to also explore the use of the knowledge base in other contexts, including an ERMS, a usage statistics manager, and its flagship EBSCONet subscription management service.

Pesch describes how EBSCO creates a mapping between its subscription management service and its knowledge base. On the subscription side, customers place orders that are tracked along with information about payment and coverage entitlements. This information can then be ported to the customer's implementation of the knowledge base, where EBSCO knows which holdings to activate based on subscription data. The value the subscription agent can bring to this exchange is the knowledge of the purchase details—whether a resource is owned or leased, bundled with other titles, and billed at the title or collection level.

That knowledge can then be used to manage additional types of data, for example, usage statistics. EBSCO's Usage Consolidation product allows libraries to import their COUNTER-compliant statistics and associates them with the appropriate holding in the

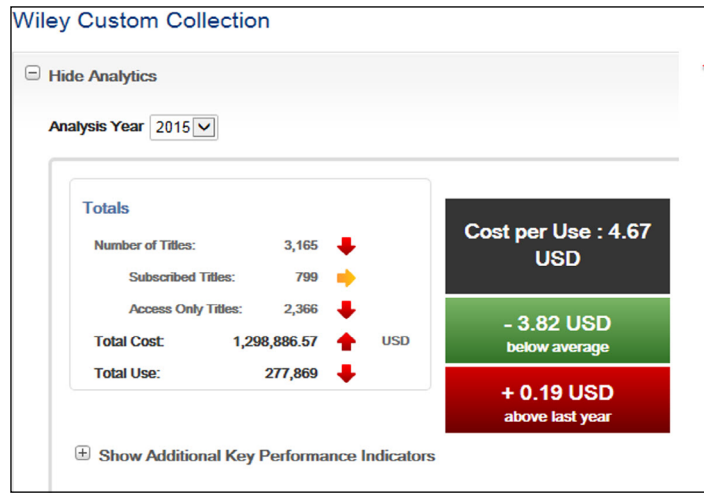
knowledge base. Associating the usage data with the holding rather than the title is essential for creating sophisticated metrics. Large libraries often purchase access to the same title through multiple sources, including traditional subscriptions, back files, and aggregated databases. Each of these purchases will accrue its own usage statistics that need to be correctly matched with payment information to calculate the popular metrics like cost per use. Usage Consolidation aids in this process by pulling together usage statistics with cost information from EBSCONet. The knowledge base holding acts as a connector between these two components, allowing the correct match to be made (see Figure 3.1).

The connection between subscription and knowledge base data has also allowed EBSCO to automate processes related to license management. On the subscription side, EBSCO receives license details from the vendors it works with. This data can be pushed into EBSCO's ERMS, ERM Essentials, where supply terms are automatically populated for an institution's collection. Pesch estimates EBSCO has been able to populate more than one million license data elements for its customers.

Dana Taylor is the head of collection management at Louisiana State University, where she uses EBSCO's subscription services as well as ERM Essentials and Usage Consolidation. Taylor said that the major benefit of the integrated products is that work can be done in a single place and then be pushed out across the EBSCO ecosystem. For LSU, orders are placed using EBSCO's subscription management tools. EBSCO can then automatically activate the purchased titles in the knowledge base, making the new resources available through LinkSource and EDS and prompting delivery of a MARC record to LSU's Sirsi ILS.

"The fact that we have that single knowledge base is extremely helpful to us," Taylor said. "Not only do the resources that we purchase from EBSCO populate our knowledge base and link resolver, but it allows us to check off resources that we've published from another publisher or vendor. It creates a seamless way to manage our entire collection."<sup>2</sup>

While EBSCO has made significant progress in bringing together discovery and electronic resources management components in its environments, the company currently lacks the ILS component that would bring in functions like acquisitions, cataloging, and circulation. As of early 2016, EBSCO has begun to publicly share its intentions to fund the development of an open-source library services platform in collaboration with the Danish software development firm Index Data and the Open Library Environment (OLE), an existing open-source LSP project. The new product,



**Figure 3.1** EBSCO's Usage Consolidation brings together cost and usage meta-data by linking them through a central knowledge base record.

called FOLIO, is expected to be engineered to promote community engagement by using a modular approach and will be built around a central knowledge base.<sup>3</sup>

## OCLC's WorldShare Management Services

OCLC's WorldShare Management Services (WMS) provides another solid demonstration of the ways that a knowledge base can unify a suite of services that covers the entire resource management life cycle. WMS includes core acquisitions, cataloging, and circulation functionality, in addition to integration with WorldCat Discovery and OCLC's resource-sharing tools. The WorldCat Knowledge Base underpins all of these services.

Jonathan Blackburn, OCLC's product analyst for WMS Acquisitions, describes how the knowledge base supports acquisitions work in WMS. To begin the ordering process, users can search the knowledge base through a widget embedded directly into the acquisitions module. This search brings up familiar knowledge base results: packages that can be expanded to show individual titles. "Add to Order" buttons embedded throughout the search results allow users to add whole packages or individual titles directly to a purchase order (see Figure 3.2). The beauty of this setup is that the purchase order itself now contains a link to the knowledge base record for the package or title that has been purchased.

"The assumption we came to this with," Blackburn said, "is that the items in the knowledge base, whether they're titles or collections, should function just the same as anything you're purchasing. It's a line item that's associated with this specific order. What's



Collection	Provider	Created	In My Holdings	Action
EBSCO MasterFILE Complete Customizable	EBSCOhost	7/23/2012, 4:59:56 PM	Held by my library (177,307) Titles Deselect Collection	Add to Order

Results 1 - 1 of 1

**Figure 3.2**  
OCLC's WorldShare Acquisitions module allows users to search for knowledge base collections and add them as line items on a purchase order.

nice, though, is that the knowledge base contains so much richer information in a lot of cases than WorldCat does, because it's got the platform and provider, the specific collection the title is a part of—all of that is being tracked.<sup>74</sup>

WMS also supports a number of other management features that rely on the integrated knowledge base. Steve McCann, the product manager for WorldShare License Manager, described how the license tool functions as a rights and responsibilities engine that associates permitted uses of a resource with the corresponding knowledge base collection. ILLiad and other ILL products can query the knowledge base, get a list of the libraries that hold the resource being requested, and ask License Manager whether or not those holdings can be shared.<sup>5</sup>

Karl Stutzman, assistant director for digital libraries services at Anabaptist Mennonite Biblical Seminary, uses the WMS suite at his small library of two-and-a-half full-time staff. Stutzman singled out the support that WMS can provide for management of demand-driven acquisitions (DDA) as another advantage of a knowledge base powered system. DDA programs allow libraries to provide their users with records for a pool of e-books that they have not yet purchased. Users trigger purchases of these books when they view a certain number of pages or meet other criteria. The way WMS works, the DDA vendor can tell OCLC what specific titles a library has active in its DDA profile and send those to the knowledge base, where they are automatically activated for discovery. From Stutzman's end, the process requires no manual intervention.

"We're in a very small library," Stutzman said. "Our ability to do something like a DDA program was nonexistent before we got this software. It really opened up the possibility, because we wouldn't have had the time and expertise to do complex record loading. So that kind of integration is really important for us."<sup>76</sup>

## ProQuest's Intota and Alma

ProQuest announced the development of a library services platform in 2011 and released the first version of Intota in 2013.<sup>7</sup> On paper, Intota has embraced the LSP vision, promising functionality to support

acquisitions, cataloging, discovery, fulfillment, and assessment with a focus on streamlined workflows and interoperability. The knowledge base was to have been the foundation of all of these services. The full version of Intota was never completed, though some of the discovery

and assessment functionality released under the Intota name has capitalized on the integrated knowledge base. The ProQuest knowledge base can receive automated holdings updates for DDA titles and other collection types, helping libraries to track complex and changeable materials. The central role of the knowledge base throughout the system also provides users with a single point of management and eliminates the need to manually create e-resources metadata.<sup>8</sup>

ProQuest has embraced the Alma LSP in the months following its official acquisition of Ex Libris in early 2016. In a January product strategy webinar, senior staff from the two companies announced that ProQuest would cease further development of Intota, instead diverting the vision for the product into continued development of Alma, which will become the flagship LSP for ProQuest. The webinar also revealed that an enhanced ProQuest knowledge base originally intended for Intota would be rolled into Alma by the start of 2017, presumably replacing the original SFX knowledge base.<sup>9</sup>

While I was not able to speak with a representative from Ex Libris for this report, published accounts of Alma implementations reveal some of the ways that the product has used an integrated knowledge base to its advantage. Chief among them is Alma's Community Zone, a hybrid of its knowledge base and a shared pool of MARC records.<sup>10</sup> Barbara Anderson, the head of metadata and discovery at Virginia Commonwealth University Libraries, described her institution's use of the Community Zone in a presentation for the Ex Libris users group. In the past, librarians at VCU had to manage their collections in each environment separately, activating resources in SFX and loading MARC records into Aleph from multiple vendors. Using the Community Zone, they can activate a title or collection in the knowledge base, and the resource will automatically become available through Primo—no need to source, store, and manage local MARC records.<sup>11</sup>

Melissa Parent and Lesa Maclean, librarians at the Royal Melbourne Institute of Technology (RMIT), addressed the impact of Alma's integrated knowledge base from the acquisitions perspective. They described Alma as inventory-centric, meaning that acquisitions records are associated with knowledge base inventory information like packages and holdings, rather than bibliographic records like in a traditional ILS.



Automated workflows can be set up that allow order records to be imported using customized profiles and linked to local inventory records. And while the initial setup for these profiles proved complicated, Parent and Maclean found that they were eventually able to automate the ordering process for most of their routine purchases. The use of inventory records as the center of the data model also brought clarity to e-resources management work by explicitly mapping relationships around a central knowledge base object.<sup>12</sup>

## Other Services

Currently EBSCO, OCLC, and ProQuest/Ex Libris are the major vendors offering products that offer a significant realization of a knowledge base integrated across a suite of services. A few other organizations have also begun to take steps in this direction, though they are not as far along.

In late 2015, Innovative Interfaces, Inc., announced that it was developing the Innovative Central Knowledge Base, which would integrate with its Sierra LSP and other services. Innovative's effort is unique among the larger vendors in that it will be acquiring EBSCO's knowledge base data to use as a foundation for the e-content portions of its knowledge base, while also leaving the door open to collaborate with other suppliers like ProQuest and Knowledge Base Plus. Innovative's arrangement will also allow mutual customers with EBSCO and other partners to seamlessly share their local holdings data between the two services.

Innovative's Central Knowledge Base will be released as a beta in the first quarter of 2016. The first version of the service will support holdings management as part of Sierra. The next release version, scheduled for the second quarter of 2016, will include integrated MARC records from Innovative's SkyRiver database, eliminating the need for users to import local versions of these records. Also scheduled for this release is a unique service that will perform automated access checking of library holdings, verifying URLs, local entitlements, and proxy configurations.<sup>13</sup>

Another open-source product, TIND, is also hoping to leverage the network effect encouraged by the community knowledge base movement. TIND spun off from Invenio, a digital asset management system originally developed to manage research documents at CERN and now overseen by an international collective. The TIND library system currently supports cataloging, circulation, and an online catalog. TIND has begun development of a combined tool to support acquisitions and e-resources management. It will be a stand-alone system that can be used with the existing TIND platform or another ILS. The system will also integrate with the GOKb open knowledge base. TIND will benefit from the work and user base that has

already begun to develop around GOKb, and its reuse of the data should also help to bring new users and contributors to that community. TIND has been working with the library community to develop its data model and services and is planning to release a beta version of its new product in late 2016.<sup>14</sup>

## Conclusion

The central role of the knowledge base in the new generation of library services platforms has already led to greater efficiency in the management of electronic resources. By serving as an identity broker, the knowledge base allows users to bring together data in intuitive and useful ways: managing purchasing alongside public activation workflows, matching up usage statistics with payments to create cost-per-use metrics, and communicating permitted uses from license records to resource-sharing systems to facilitate unmediated lending. While it's hard to quantify the exact labor savings for libraries, the examples provided by smaller libraries show that LSPs are making complex programs like DDA viable when they would not have been possible using traditional systems.

Undoubtedly, more features will become available as these tools continue to evolve. Areas of potential innovation include workflow management tools that guide users through the resource life cycle and automate tasks when possible; round-trip communication, allowing users to make changes anywhere in the system and push them out to the knowledge base; and increased reliance on knowledge base-like environments for other types of data, such as MARC records, article-level metadata, and instructional materials. With the promise of the integrated knowledge base beginning to crystalize, creative organizations have the opportunity to drive a real evolution in library systems technology.

## Notes

1. Oliver Pesch (chief strategist at EBSCO) in discussion with the author, November 2015.
2. Dana Taylor (head of collection management at Louisiana State University) in discussion with the author, December 2015.
3. Sebastian Hammer, "Constructive Disintegration—Re-imagining the Library Platform as Microservices," (presentation, Code4Lib 2016 conference, Philadelphia, PA, March 9, 2016).
4. Jonathan Blackburn (product analyst for WorldShare Management Services Acquisitions at OCLC) in discussion with the author, November 2015.
5. Steve McCann (product manager for WorldShare License Manager at OCLC) in discussion with the author, November 2015.
6. Karl Stutzman (assistant director for digital library services at Anabaptist Mennonite Biblical Seminary)

- in discussion with the author, December 2015.
7. Marshall Breeding, "Library Services Platforms: A Maturing Genre of Products," *Library Technology Reports* 51, no. 4 (May/June 2015): 26–36, <https://journals.ala.org/ltr/issue/view/509>.
  8. ProQuest, *Intota and Our Knowledgebase: Global Metadata from a Central Source*, datasheet, accessed April 29, 2016, <http://media2.proquest.com/documents/+D3935+-+Intota+Knowledgebase+Datasheet.pdf>.
  9. "Ex Libris and ProQuest Product Strategy and Roadmap," webinar, ProQuest, January 13, 2016.
  10. Kristen Wilson, "Introducing the Next Generation of Library Management Systems," *Serials Review* 38, no. 2 (2012): 110–23, <http://dx.doi.org/10.1080/00987913.2012.10765438>.
  11. Barbara Anderson, "The Parallel Universe of Electronic Resources Discovery, or How I Learned to Stop Worrying and Love the Community Zone" (presentation, Ex Libris Users of North America Annual Meeting, Minneapolis, MN, May 5–8, 2015), [http://scholarscompass.vcu.edu/libraries\\_present/34](http://scholarscompass.vcu.edu/libraries_present/34).
  12. Melissa Parent and Lesa Maclean, "Go with the Flow: Discovering New Workflows in Alma" (presentation, VALA conference, Melbourne, Australia, February 3–6, 2014), VALA2014 Proceedings website, [www.vala.org.au/vala2014-proceedings/vala2014-session-14-parent](http://www.vala.org.au/vala2014-proceedings/vala2014-session-14-parent).
  13. Tim Auger (director of resource sharing and e-content at Innovative Interfaces, Inc.) in discussion with the author, March 2016.
  14. Kristen Wilson, "Building the Global Open Knowledgebase (GOKb)," *Serials Review* 39, no. 4 (2013): 261–65, <http://dx.doi.org/10.1080/00987913.2013.10766408>.

# New Uses for Knowledge Bases

The usefulness and ubiquity of knowledge base data in large research and academic libraries has led to much innovation in this space. While knowledge bases have evolved continuously since their introduction, the past five years have seen increased engagement with these tools across the field, leading to a number of exciting developments. These include new thinking about how knowledge bases are structured and the data they collect, increased use of APIs to integrate knowledge bases with new services, and trends toward greater automation and customization.

## Enhanced Knowledge Base Data

The type of data found in a knowledge base has remained fairly consistent since the earliest implementations. Titles, packages, and holdings, along with their associated attributes, remain the core data elements. Recently, however, some knowledge base suppliers have begun to rethink the basics and explore enhanced data models for their knowledge bases. Two major efforts in this area include the development of a re-architected knowledge base by ProQuest and exploration of an enhanced, librarian-driven data model by the Global Open Knowledgebase (GOKb).

## ProQuest's New Knowledge Base

In late 2015 ProQuest announced a new knowledge base designed to enrich its existing service, which has its roots in the original Serials Solutions knowledge base dating back to 2001.<sup>1</sup> Yvette Diven, product manager lead for management solutions at ProQuest, described work in four key areas for the company's knowledge base: scope, scale, systems, and services.

The scope of the ProQuest knowledge base will become more global and diverse through the inclusion of new electronic content types, including streaming audio and video titles. The new knowledge base also will feature a single data model that pulls together the traditional e-resources metadata, along with the contents of Ulrich's Periodicals Directory, authoritative information from MARC records, and article-level metadata from the Summon discovery index. The reengineered product will live in the cloud, making it more scalable. And an API will allow ProQuest to reuse this enriched metadata across all of its products and services and to share the data more widely with its customers.

Diven said that these changes will give ProQuest's customers a comprehensive view of their collections from within a single integrated product. The enriched knowledge base will also map the relationships between entities in a more sophisticated manner—making connections between, for example, an author and a title, an organization and the resources it publishes, and two journals published by the same entity. The knowledge base will also have the ability to track changes, helping users manage title and publisher changes and allowing them to see snapshot views of their collections over time.<sup>2</sup>

ProQuest has been gradually rolling out these changes across its existing product suite, and Diven describes the process as more of a continuum than a migration. The enhanced knowledge base was planned to be a major component of Intota, ProQuest's developing LSP. In a recent webinar, the company announced that the vision for Intota, including use of the new knowledge base, instead would be rolled into Alma, the LSP ProQuest recently took over with its acquisition of Ex Libris.<sup>3</sup>

## GOKb's Enhanced Data Model

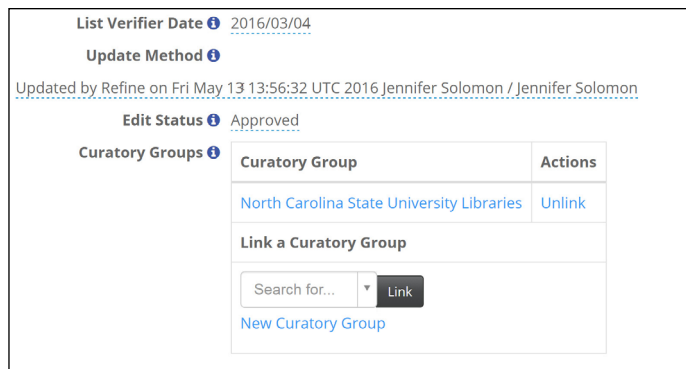
(Full disclosure: I am the principal investigator of the Global Open Knowledgebase (GOKb) project, and any uncited information regarding the project in this section comes from my personal experiences.)

GOKb is a community-managed, open-source project that aims to make e-resources metadata freely available to the library community. Like ProQuest, the GOKb has been innovating in the knowledge base space by addressing the data itself. The GOKb data model has been designed with the flexibility to model a complex environment and the transparency to work openly through a community contribution model.

The goal of creating a data model that can handle the current electronic resources landscape, as well as expand to accommodate changes in the market, led the GOKb development team to adopt the bill of materials (BOM) approach. Used widely in industry, the BOM model labels individual items as *components*, which can be bundled together into *combinations*. New component and combination types can always be created, and combinations can even be linked together to form larger combinations. In the current GOKb environment, three components—titles, packages, and platforms—are linked together to form a combination that represents a holding. But if, in the future, the knowledge base needs to accommodate, for example, article-level metadata, the model can easily accommodate this expansion. Articles can simply become a new type of component, bundled together into journals, which now become combinations.<sup>4</sup>

Like ProQuest, GOKb is working to track changes over time, including title changes and transfers between publishers. Using the BOM model, GOKb allows users to create linkages between two titles to represent a change. All of the titles linked together in this way can be pulled together to represent a comprehensive title family. Similarly, the BOM model also allows a linkage to be created between a title and the organization that publishes it. For any title, users can view all of the title-publisher linkages, along with associated dates, to see a comprehensive publication history.<sup>5</sup>

As a community-managed knowledge base, GOKb has also taken the unique step of building transparency into its data model. Since project partners from many different universities have a role in creating and maintaining data, it's important for users to be aware of who's doing what. To this end, GOKb has included fields in all of its components that allow users to see who has last updated that record. For core components like packages and titles, GOKb also includes several additional status fields. These include the name of an individual verifier, last verified date, and an approval



**Figure 4.1**

A GOKb package record contains information about the group responsible for its maintenance and the individual user who last edited it.

status (see Figure 4.1). Packages can also be assigned a curator—an institution that has claimed responsibility for managing that particular group of titles. While users outside the curator group can still edit the package, the system provides a warning message and encourages users to communicate with a curator before making major changes. Taking inspiration in part from the model used by *Wikipedia*, the goal of these fields is to encourage communication and trust between the users of GOKb.

## Knowledge Base APIs

Application programming interfaces (APIs) are sets of tools for building and interacting with software applications.<sup>6</sup> In recent years, many types of library systems and services have begun offering APIs that allow users to build their own integrations with a vended product. Knowledge bases are no exception to this trend. OCLC currently offers an API for the WorldCat Knowledge Base, and both ProQuest and Innovative are planning to introduce them soon as part of their knowledge base enhancements. These APIs are beginning to give knowledge base customers the flexibility to create custom solutions using knowledge base data.

OCLC's WorldCat Knowledge Base API is available in a sandbox version to anyone, but in production only to libraries that use the knowledge base. The API can provide article, e-journal, or e-book citations; links to e-resources customized with a user's account identifiers; proxy information; and browse and search features similar to an A-to-Z list.<sup>7</sup>

Brian Cassidy, senior web developer at the University of New Brunswick (UNB), shared some details about his library's use of the WorldCat Knowledge Base API to create a custom discovery tool. The library's website features several search tabs for different types of e-resources, including databases, journals and

newspapers, online reference works, e-books, and videos. All of the results returned by these searches are drawn straight from OCLC's knowledge base via the API. Users can search for specific titles, browse lists of collections, and link out to their desired resource (see Figure 4.2).

Very little of this functionality came out of the box, Cassidy said, but rather was all designed in-house by staff at UNB. To use the API, UNB's system provides OCLC's API with a web service key that authenticates it as a valid user and authorizes the API to release customer information. The UNB website makes web requests to the API and receives JSON or XML code in return, which it can then use to create the custom search environment. UNB is also preparing to integrate its custom search with OCLC's WorldShare License Manager, which will facilitate the display of permitted uses along with the search results.<sup>8</sup>

Since the WorldCat Knowledge Base API is fairly new, it will take time before more libraries can experiment with the functionality and discover new ways of using it. Stephanie Doellinger, section manager for data services at OCLC, said that creating homegrown A-to-Z lists and search interfaces continues to be the most popular use of the service at this time. Jodie Stroh, OCLC's product manager for Collection Manager, suggested that the API could also potentially be used to expose a library's unique digitized collections. A library could create a custom collection in the knowledge base with links to archives, photographs, or videos. That metadata would then be available to other knowledge base users to expose through implementations of OCLC services.<sup>9</sup>

## Integrations with Other Products

Knowledge bases have also proved a practical way to communicate information about an institution's collection to outside services. Most academic libraries already use knowledge bases to support core discovery and management tools—usually all centralized with a single large vendor. But the same holdings information stored in a knowledge base is often required by other services as well. Rather than duplicate the effort of describing the same collections information in two (or more!) places, librarians are working to find creative ways to reuse their knowledge base metadata to help support a broader array of products.

Steve Oberg, assistant professor of library science, described how Wheaton College, in conjunction with the CARLI consortium, has been using holdings information pulled from the SFX knowledge base to support its implementation of BrowZine, a browsable interface for scholarly journals. For the service to work correctly,

The screenshot shows a search interface with a navigation bar at the top containing 'All', 'Article Databases', 'Journals & Newspapers', 'e-Encyclopedias, etc.', 'e-Books', and 'Video'. Below the navigation bar, a message states: 'Your search for "tetrahedron" found 5 records. Would you like to Q search again?'. The search results are displayed in a table with three entries:

<b>Tetrahedron</b> Science Direct v. 1 1957-Current	ISSN: 0040-4020 WorldCat: 39263349
<b>Tetrahedron: Asymmetry</b> Science Direct v. 1 1990-Current	ISSN: 0957-4166 WorldCat: 39263360
<b>Tetrahedron Computer Methodology</b> Science Direct v. 1 1988-v. 3 1990	ISSN: 0898-5529 WorldCat: 63800493

**Figure 4.2**

The University of New Brunswick Libraries' custom search interface is generated using the WorldCat Knowledge Base API and contains direct links into WorldCat.

customers need to communicate to BrowZine exactly which journals their library subscribes to. And while BrowZine provides a way to manually input local holdings, doing so would duplicate work that consortium members have already done in their shared implementation of SFX. With this in mind, Oberg and his colleagues began working with Ex Libris to create a solution that would allow them to use their knowledge base holdings to communicate with BrowZine.<sup>10</sup>

The resulting process involves procuring a weekly export of all active full-text holdings from SFX, which is output to a zip file and stored in an accessible directory on the SFX server. BrowZine then fetches that file and uses it to rebuild each library's holdings information. This system builds on existing SFX functionality that allows customers to set up export profiles based on locally defined criteria. Oberg said that CARLI was one of the first BrowZine users to implement the automated system and the first consortial user to do it. Since then, BrowZine has expanded this functionality to work with other knowledge bases and makes the process available as a standard part of its service.<sup>11</sup>

The experience of integrating BrowZine with the SFX knowledge base has prompted Oberg and the CARLI SFX committee to pursue a new research project that will explore ways to make use of the SFX data with other services, including possibilities like WorldCat Local, Google Scholar, and ILLiad.

"It's not just an SFX thing," Oberg said. "It's something people need to think about a lot more. It's how to leverage all of your investment in your knowledge base, to reduce duplicate work and make sure your access is consistent. You want to make sure that whatever path users choose, they're able to get to your resources. To me the knowledge base is the key part there."<sup>12</sup>

## Delivery of Library-Specific Holdings

Another trend that highlights the importance of efficiency and accuracy of knowledge base data is the

move toward delivery of library-specific holdings directly to knowledge bases. Up until now the supply chain has primarily focused on the delivery of global information from publisher to vendor, with the library supplying the localization component. But movement in this space suggests this equation may soon change. Since publishers must keep track of their customers' holdings to manage access and billing, it makes sense that these publishers could also communicate those holdings directly to knowledge base suppliers on behalf of their customers.

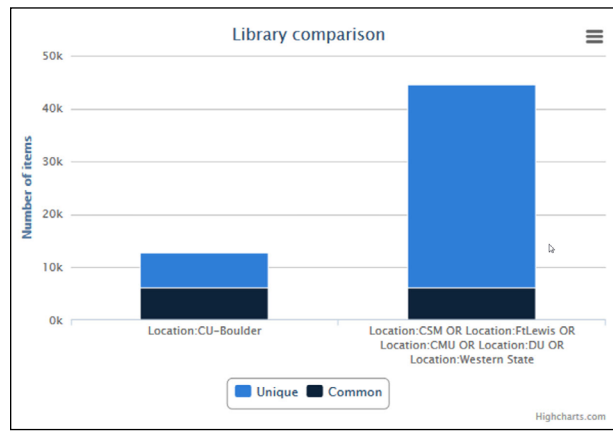
The knowledge base providers I spoke with agreed that library-specific holdings would continue to be a key area of expansion in the knowledge base space. Stephanie Doellinger said that OCLC's customers love their existing vendor feeds with Elsevier, Ebrary, and EBL, and that the addition of new feeds is one of their top requests.<sup>13</sup> Oliver Pesch from EBSCO echoed the importance of custom feeds, but stressed that additional functionality would be necessary to make them work for management as well as discovery.<sup>14</sup>

Pesch has been working with Elsevier and others throughout the supply chain to submit a proposal to NISO for work that would enhance the KBART best practice with new functionality to help support delivery of customized holdings, in addition to general efficiency improvements. The proposal notes that libraries care about management metadata and specifies the need to develop a best practice for delivering feeds that include both entitlement and packaging information. It also includes work in the area of automating delivery of data from publisher to knowledge base supplier (for both global and custom feeds) using a web service. The proposal has been endorsed by EBSCO, GOKb, Elsevier, and Project COUNTER.<sup>15</sup>

## Decision Support

On the management front, knowledge bases are beginning to be thought of as tools to aid in what has come to be known as *decision support*—the process of gathering information to help with selection and ongoing maintenance of e-resources. Gold Rush, an electronic resources management system and knowledge base offered by the Colorado Alliance of Research Libraries, has made a niche for itself in the decision management space. Gold Rush is a smaller nonprofit service, and many of its customers also subscribe to the larger vended discovery services, said George Machovec, the Alliance's executive director. Rather than try to compete in this arena, Gold Rush has focused on developing a suite of services that allow its customers to delve more deeply into the analytics space.<sup>16</sup>

Gold Rush Decision Support allows libraries to compare packages and to determine the unique and common titles between the two. The service supports



**Figure 4.3** The Gold Rush Library Content Comparison System expands the concept of a traditional knowledge base overlap analysis tool to include MARC records and to compare holdings across multiple libraries.

full-text packages, as well as indexing and abstracting services. Machovec said that these features help libraries make better decisions about what to purchase, but can also support maintenance activities in unique ways. For example, the Decision Support Tool can be used to compare a library's holdings in a publisher collection to the holdings found in the Portico or CLOCKSS packages to help investigate compliance with archiving best practices.

A unique feature of Gold Rush is the Library Content Comparison System, which allows libraries to upload their MARC records to a knowledge base-like space (see Figure 4.3). This service is particularly useful if multiple libraries within a consortium subscribe to the service. Participants can then compare their MARC holdings against their peers using a matching algorithm based on various pieces of data, including the title, publisher, and fixed fields. It's a clever repurposing of the knowledge base concept to support different types of collection needs.

Decision support has also become key for e-books, which can be purchased on a dizzying array of platforms, subject to complex technical limitations, and sold in unwieldy bundles of thousands of titles. A 2014 report from the Jisc e-book Co-Design Project aimed to understand the pain points surrounding e-book management and propose actions to address these issues.<sup>17</sup> The top pain points identified in the report included finding out what e-books are available, providing continuing and archival access to purchases, and managing e-book usage statistics.

Jisc's involvement with Knowledge Base Plus (KB+) and GOKb led to two recommendations for development in these systems to advance decision support. The first is advanced availability tracking, which aims to normalize e-book metadata and

identifiers and to track movement of e-books in and out of packages. The second, focused on decision support, proposes a data model that would allow libraries to contribute practical information about e-book management to the global knowledge base. Examples of the type of data to be tracked include license terms, formats offered, digital rights management (DRM) restrictions, platform characteristics, and device compatibility. Some universities in the United Kingdom are already tracking this information using spreadsheets and other local tools. If these attributes were stored in a global platform, they could reach a wider audience of users who could both benefit from and contribute to the decision support data. Since the publication of the Co-Design report, GOKb (with support from Jisc), has begun building a prototype for the e-book availability tracking and decision support functionality. A production release is scheduled for the fall of 2016.

## Conclusion

The examples in this chapter demonstrate that the knowledge base has broken through the boundaries of core technical services work to become a key data repository that intersects with workflows across the library. New types of information are being added to knowledge bases that will allow users to manage streaming media formats, track changes over time, and automatically receive customized updates to their holdings. Export processes and APIs allow libraries to use their knowledge base data in more contexts than ever, supporting streamlined management of multiple tools and allowing for the creation of custom interfaces. And more sophisticated data means that knowledge bases can be used for new decision support purposes like availability tracking and collection analysis.

Even with all of this progress, more possibilities remain to be explored. The knowledge bases of the future may allow libraries to implement unmediated borrowing and purchasing at a greater scale through integrations with document delivery services. New data output formats, especially linked data, may support more fluid communication with external systems. And increased need for data to be open and reusable across multiple systems and services may improve interoperability and even lead to the adoption of more central, nonproprietary knowledge base solutions. Chapter 5 of this report will explore the beginnings of this last concept in greater detail.

## Notes

1. Yvette Diven, "What's This I Hear about a New Knowledgebase? Part 1, Why a New KB?" *ProQuest Blog*, December 14, 2015, [www.proquest.com/blog/pqblog/2015/Whats-this-I-Hear-about-a-New-Knowledgebase-Part-1-Why-a-New-KB-.html](http://www.proquest.com/blog/pqblog/2015/Whats-this-I-Hear-about-a-New-Knowledgebase-Part-1-Why-a-New-KB-.html).
2. Yvette Diven (product manager lead for management solutions at ProQuest) in discussion with the author, October 2015.
3. "Ex Libris and ProQuest Product Strategy and Roadmap," webinar, ProQuest, January 13, 2016.
4. Kristen Wilson, "Building the Global Open Knowledgebase (GOKb)," *Serials Review* 39, no. 4 (2013): 261–65, <http://dx.doi.org/10.1080/00987913.2013.10766408>.
5. Kristen Wilson, "Bringing GOKb to Life: Data, Integrations, and Development," in *The Importance of Being Earnest: Charleston Conference Proceedings, 2014*, ed. Beth R. Bernhardt, Leah H. Hinds, and Katina P. Strauch (West Lafayette, IN: Purdue University Press, 2015), 607–13, <http://doi.org/10.5703/1288284315649>.
6. *Wikipedia*, s.v., "Application programming interface," accessed February 20, 2016, [https://en.wikipedia.org/wiki/Application\\_programming\\_interface](https://en.wikipedia.org/wiki/Application_programming_interface).
7. "WorldCat Knowledge Base API," OCLC Developer Network, <https://www.oclc.org/developer/develop/web-services/worldcat-knowledge-base-api.en.html>.
8. Brian Cassidy (senior web developer at the University of New Brunswick) in discussion with author, November 2015.
9. Stephanie Doellinger (section manager for data services at OCLC) and Jodie Stroh (product manager for Collection Manager at OCLC) in discussion with author, October 2015.
10. Steve Oberg (assistant professor of library science at Wheaton College) in discussion with author, November 2015.
11. "Working with Your Library Holdings," knowledge base, Third Iron, <http://support.thirdiron.com/knowledgebase/topics/79846-holdings-reports-information>.
12. Steve Oberg (assistant professor of library science at Wheaton College) in discussion with author, November 2015.
13. Doellinger discussion.
14. Oliver Pesch (chief strategist at EBSCO) in discussion with the author, November 2015.
15. Oliver Pesch, e-mail message to author, November 19, 2015.
16. George Machovec (executive director of the Colorado Alliance of Research Libraries) in discussion with author, November 2015.
17. Ian Chowcat, David Kay, and Owen Stephens working with Amy Devenney and Graham Stone, *eBooks Co-Design Report* (London: Jisc Collections, March 31, 2014), [www.jisc-collections.ac.uk/Global/Projects/KB+/KB+%20Documents/140423-ebooks-co-design-report-final.pdf](http://www.jisc-collections.ac.uk/Global/Projects/KB+/KB+%20Documents/140423-ebooks-co-design-report-final.pdf).



# Collaborative Knowledge Bases

The idea of an open, central, and collaboratively managed knowledge base is as old as the knowledge base itself. The first project of this type was the Jointly Administered Knowledge Environment (jake), which began at Yale University in 1999. The goal of the project was to track e-resources metadata and relationships in an open-source environment. Librarians with an interest in the project were encouraged to contribute by collecting journal title lists, correcting errors, and promoting the project with publishers and vendors. By banding together, jake participants could help reduce the duplication of effort that occurred when individual libraries each had to research and document the same information about e-journals.<sup>1</sup> While jake shut down for good in 2007 and never existed as more than a simple online reference of e-resources metadata, it helped set the stage for future efforts to develop open community knowledge bases.

Culling engaged in a significant discussion of the centralized knowledge base in his 2007 report to UKSG. He pointed out that vendors, like librarians, also engage in duplication of effort when it comes to managing e-resources metadata. Each knowledge base supplier must build and maintain its proprietary product in isolation—even though these products all strive to describe the exact same universe of resources. He proposed as an alternative a single central knowledge base that would use web services to provide its data freely to anyone who wished to use it. Culling concluded that while a centralized solution might be possible in the long-term future, it would require significant investment and management from an organization that had the resources to support it.<sup>2</sup>

Another eloquent plea for a centralized knowledge base came from Singer in a 2008 article. He disputed the notion that a single entity would need to manage

such a knowledge base and instead pointed to successful projects like *Wikipedia* and the Internet Archive, which harness the power of many invested users to manage open, dynamic content. Singer acknowledged the difficulties of creating such a service, including modeling complex data and coordinating the involvement of large numbers of data managers. However, he believed the payoff in implementing this model would ultimately be worth the cost:

The knowledgebase crisis is not going away, and as the digital universe expands, especially to new and different formats, it will only get more difficult to manage. By tapping into the power of the entire community—from the beginning of the publishing chain to the end-user—the knowledgebase becomes self-sustaining and finds new and interesting uses along the way.<sup>3</sup>

While Singer's vision has certainly not become reality yet, several projects that have emerged over the past five years demonstrate that the desire remains to collectively improve knowledge base data and ease its flow across the supply chain.

## Community-Managed Knowledge Bases

### The Global Open Knowledgebase

(Full disclosure: I am the principal investigator of the GOKb project, and any uncited information regarding the project in this section comes from my personal experiences.)

The project most closely aligned with the grand vision for knowledge base collaboration is the Global



Open Knowledgebase (GOKb.) Not unlike Jisc, the project aims to provide a fully open, community-managed knowledge base that describes electronic journals and books and their relationships. The three major ambitions for the GOKb project are improving data quality across the supply chain, reducing duplication of effort, and encouraging interoperability between systems. GOKb's focus on openness, collective effort, and enhanced data model (described in chapter 4) all contribute to its work in these areas.

The GOKb project began as a joint venture between Jisc and the Quali OLE project. In addition to support provided by these institutional project partners, GOKb also employs one full-time staff member, the GOKb editor. The editor is responsible for setting the policies that define how the data is managed and for coordinating the community members who can contribute various forms of effort to GOKb. Contributions include collecting and loading KBART-formatted title lists into the knowledge base, addressing data errors and anomalies identified during the loading process, and engaging in other data enhancement activities, such as researching and documenting title history information.<sup>4</sup> As the lead school on the project, North Carolina State University has engaged heavily with GOKb, contributing staff time to pilot a data-loading initiative, and several other Quali OLE partners have contributed to the data-loading process as well.<sup>5</sup> GOKb has also been successful in attracting librarians unaffiliated with its major partner projects to work with the knowledge base in more lightweight ways—particularly in areas such as researching title changes and documenting them in the knowledge base.

GOKb's data is freely available under a Creative Commons 0 (CC0) license, which means that it can be used by anyone, for any purpose, without attribution.<sup>6</sup> While GOKb was originally created to support the Knowledge Base Plus (KB+) and Quali OLE services, the fact that the data is in the public domain means the project can have a much broader impact. Other open-source projects in need of knowledge base data are free to use GOKb, and—just as importantly—publishers and vendors can consume the data as well. As GOKb grows and the quality of its data improves, publishers at the top of the supply chain can also use GOKb's data to improve their own data, while knowledge base suppliers can integrate the data into their services. Vended knowledge bases could at some point even replace their proprietary knowledge bases with GOKb or mirror some or all of its content rather than maintaining the same information themselves.

The visionary changes to which GOKb aspires are still a ways off. GOKb currently includes about 400 packages, compared to the tens of thousands found in most commercial systems. The scale of data that a comprehensive knowledge base needs to cover has proven difficult to achieve with only a single staff member

and a couple dozen volunteers. The development team for the project is currently working on a new data loader that will allow multiple files to be loaded at once, opening up the possibility of consuming larger data sources in an automated way. New partners will also be necessary to achieve scale. Library partners are needed to help monitor data quality and collect enhanced data like title and publisher changes. And a large-scale partner—possibly even another knowledge base—will likely be required to collect the amount of data needed to be truly comprehensive.

Still, GOKb exists as an excellent proof of concept of the open, collaborative knowledge base. My experiences working with this project have convinced me that the library community values work in this area and that many individual librarians would be willing to contribute to an easy-to-use, well-managed knowledge base effort. I believe, also, that buy-in from other stakeholders, including publishers, knowledge base vendors, and standards organizations, is essential to meeting this goal. The vision for a centralized knowledge base remains valid, but it cannot be fully realized without the engagement of key players across the supply chain.

### WorldCat Knowledge Base

OCLC has also begun to explore a community management approach with its WorldCat Knowledge Base. While this product is not open source or intended to be a cross-product solution, OCLC has gone further than any of the other vended knowledge base products in inviting librarians to be part of the management process.

The WorldCat Knowledge Base operates using a cooperative approach that allows customers to view changes made to the knowledge base and vote on whether to approve or deny them. The voting window for each change is open for five days. If a change gets ten votes in either direction during this window, it will be implemented or rejected accordingly. If fewer than ten votes are received, the change will be automatically accepted when the voting window closes. Jackie Fahmy from OCLC said that users tend to cast more negative votes for errors and problematic changes, and simply let the voting window expire for the changes that don't affect them.<sup>7</sup> Votes also tend to come from a small number of very active libraries that want a lot of control over their data. To increase participation, OCLC has considered implementing a notification service, which would allow users to receive alerts when changes occur in specifically chosen packages.

OCLC also offers its users the ability to create custom packages that can be shared globally with all of its knowledge base customers. In addition to supplying typical knowledge base data, users can also



**Figure 5.1**

The BACON knowledge base helps French publishers improve their metadata and assigns a quality label to those who meet certain standards.

link up the holdings in custom collections with the appropriate MARC records. Fahmy said that creating custom collections for packages where the publisher doesn't provide KBART files or MARC records is a popular use case. Participants in consortial deals have also taken advantage of the cooperative management functionality. Fahmy described how one North Carolina library created custom collections to represent some of the content it receives from NCLIVE, a statewide consortium. The packages were then made available to other OCLC libraries in North Carolina that had access to the same content. In this way, individual librarians, many of whom are doing knowledge base work anyway, can have an impact beyond just their own institutions.

"Being a cooperative ourselves here at OCLC, we thought it was a good idea to allow our knowledge base to be a cooperative, too," Fahmy said. "We're reliant on data from providers and knowing that not everything is perfect, we wanted to give users the ability to make data changes for everybody. By doing this, we're giving the cooperative and the librarians the ability to own this data and make it what they need it to be."<sup>8</sup>

## National Knowledge Bases

National knowledge base projects have also taken up the banner of open, collaborative data management. These projects are most often run through a government agency, national library, or large university, and they attempt to create central knowledge bases describing resources specific to a certain country.

National knowledge bases have tended to emerge in countries where there is already a high level of national collaboration, including the United Kingdom, Germany, France, and Japan, among others.

One of the primary goals of national knowledge bases has been to improve the accuracy of data that commercial suppliers often struggle to provide. National knowledge bases tend to fall into two categories with regard to this goal. Some aim to describe electronic resource content purchased by libraries in their country, regardless of its origin. The KB+ project in the United Kingdom, for example, describes subscription deals negotiated by British consortia, along with some master title lists for popular publishers. KB+ data managers spend huge amounts of time verifying title lists and improving metadata. Earney noted that KB+ data managers spent more than 70 hours creating a single title list for at least one major publisher package.<sup>9</sup>

In Germany, two knowledge base-like projects also attempt to capture definitive lists of holdings on behalf of member libraries. The Zeitschriftendatenbank (ZDB), or German Journal Database, is a bibliographic database that contains MARC records representing the print and online journal collections held by more than 4,400 German and Austrian libraries.<sup>10</sup> The Elektronische Zeitschriftenbibliothek (EZB), or Electronic Journals Library, provides information about German-held online serials, with more of an electronic resources management perspective.<sup>11</sup> In both cases, small, dedicated staffs collect and vet information with the goal of providing highly accurate metadata.

Other national knowledge bases focus more on describing publications that originate from their homelands. In France, the BAse de CONnaissance Nationale (BACON) project has a mission to create high-quality knowledge base data describing French publications.<sup>12</sup> The idea for BACON came up when the ABES agency, which maintains a French union catalog, conducted a survey that revealed that most libraries were happy with their vended tools, but found that data about French publications was often missing or incorrect. BACON aims to close this gap by collecting title lists from French publishers, analyzing and correcting errors, and formatting the lists according to the KBART code of practice. The vetted lists are then made freely available through the BACON site and shared with the original publishers (see Figure 5.1). The E-Resources Database-Japan (ERDB-JP) is a similar effort to supplement the supply chain with knowledge base data describing electronic journals and books written, edited, or published in Japanese.<sup>13</sup>

The data created by national knowledge bases is designed to directly benefit constituent libraries. In some cases, the data serves mainly as a reference. It can be searched and browsed on the web, exported for

local use, or accessed via an API. Benjamin Bober, the manager of the BACON project, described one potential use case for the French knowledge base data. ABES has been working on developing a tool to analyze e-resource usage using EZProxy logs. The goal of the project is to pull each URL visited from the logs and determine which resource it corresponds to using knowledge base data. This data can then be compared with COUNTER usage data to confirm accuracy or used in place of COUNTER data when it is not available. Without accurate, KBART-compliant files for French publications, such a process would be all but impossible.<sup>14</sup>

Other national knowledge bases go a step further by building services on top of their data. The ZDB and EZB, both of which have been around since 1997, support core library services. The ZDB provides tools to facilitate interlibrary loan and document delivery, and the EZB offers a linking service and XML feeds that can be used to support integration of the data with other systems. KB+ powers a full-featured electronic resources management system (ERMS) that supports subscription management, licensing, and integration with the Journal Usage Statistics Portal (JUSP). ERDB-JP supports a link resolver and discovery service.

The value of national knowledge bases also extends beyond library-focused services. KB+, BACON, and ERDB-JP all have explicit goals of improving the supply chain by making their data freely available for reuse under a CC0 license. In many cases, the knowledge base data created by these organizations fills in vital gaps in the supply chain. From the perspective of Tomoki Ueno and Tomoko Kagawa, who help manage the ERDB-JP project, Japanese resources are often underrepresented in products that are aimed at predominantly American and European audiences. By creating an open knowledge base of Japanese materials, they hope to provide a reliable source for this missing data.<sup>15</sup> The consortial deals managed by KB+ can also be difficult for commercial knowledge bases to represent because the details are not included in public channels and official data feeds. Currently, OCLC, ProQuest/Ex Libris, and EBSCO are all using freely available KB+ to enhance their own knowledge bases.

“The collaboration with KB+ has been fantastic,” Yvette Diven, product manager lead for management solutions at ProQuest, said. “The entire team there is trying to build up something that’s repeatable, that’s sharable for other groups. And they’re very open to working with commercial knowledge bases. They’ve actually laid the groundwork so that others who are using the KB+ model can follow. They’ve broken that ground.”<sup>16</sup>

Efforts to mend the supply chain extend all the way back to the source, as many national knowledge base

creators also make it their goal to work directly with publishers. Such work is an explicit part of BACON’s mission. The knowledge base assigns a special quality label to identify data that has been certified by ABES and adopted by the publisher to improve the data provided on its own platform. So far, only two publishers have earned the quality label, but Bober said that BACON is working closely with nine or ten other publishers and hopes to award more endorsements soon. “It’s a win-win because we centralize the metadata, but also encourage publishers to use it on their own platforms,” he said. “So far, I think publishers are quite happy with the work we have done.”<sup>17</sup>

GOKb has been in contact with all of these national knowledge base representatives to discuss the logistics of incorporating their data into the global knowledge base. The open nature of the data means the only barriers to this type of collaboration are the technical and resource challenges of loading and updating large amounts of data. Enhancements to GOKb’s data-loading process should make redistribution of this data a more realistic prospect before 2016 is out.

## Conclusion

Enormous political and structural challenges stand in the way of fully implementing the open knowledge base vision. Purveyors of commercial products view the quality of their knowledge bases as a sales differentiator and would be rightly cautious in abandoning their proprietary systems for a communal approach. Individual libraries and librarians are often stretched thin and may believe that a vendor with paid staff could simply do the work better. Any change to the current situation will likely be a long and gradual one.

Still, in the interplay between open, commercial, and national knowledge bases and their users, it’s possible to see how the vision for community knowledge base management might eventually play out. Participants in each type of knowledge base can contribute work that is natural and meaningful to their circumstances to the larger community. Commercial and global services would likely collect the data with the broadest application, national and regional groups would have an incentive to supplement it with specialized collections, and users across the board would contribute individual enhancements and corrections about the titles that are most important to them. Together these groups are already performing much of the work that would be needed to support more centralized knowledge base management across the industry.

The benefits of a truly central knowledge base to the field would be enormous. Nearly every trend I’ve written about in this report would benefit from

greater openness of knowledge base data. Shared identifiers (or even a shared data pool) could contribute to greater interoperability across tools and platforms, offering libraries choice and flexibility. Data enhancement efforts could increase, as open knowledge bases could be easily paired with other free data sets, especially those emerging with the rise of linked data. Broad availability of the data through APIs would increase creativity throughout the field, allowing libraries and individuals to create their own knowledge base-powered projects. And maximizing the number of users of a single data pool could have a big impact as different types of users contribute data changes and enhancements that can be applied across the entire supply chain. While achieving this vision will be no easy feat, the potential for great strides exists, and the first steps have already been taken.

## Notes

1. Daniel Chudnov, Cynthia Crooker, and Kimberly Parker, "jake: Overview and Status Report," *Serials Review* 26, no. 4 (2000): 12–17, <http://dx.doi.org/10.1080/00987913.2000.10764619>.
2. James Culling, *Link Resolvers and the Serials Supply Chain*, final project report for UKSG (Oxford, UK: Scholarly Information Strategies, May 21, 2007), [www.uksg.org/projects/linkfinal](http://www.uksg.org/projects/linkfinal).
3. Ross Singer, "The Knowledgebase Kibbutz," *Journal of Electronic Resources Librarianship* 20, no. 2 (2008): 85, <http://dx.doi.org/10.1080/19411260802272776>.
4. Kristen Wilson, "Bringing GOKb to Life: Data, Integrations, and Development," in *The Importance of Being Earnest: Charleston Conference Proceedings, 2014*, ed. Beth R. Bernhardt, Leah H. Hinds, and Katina P. Strauch (West Lafayette, IN: Purdue University Press, 2015), 607–13, <http://dx.doi.org/10.5703/1288284315649>.
5. Eric M. Hanson, Xiaoyan Song, and Kristen Wilson, "Managing Serials Data as a Community: Partnering with the Global Open Knowledgebase (GOKb)," *Serials Review* 41, no. 3 (2015): 146–52, <http://dx.doi.org/10.1080/00987913.2015.1064853>.
6. "CC0 1.0 Universal Public Domain Dedication," Creative Commons, accessed April 29, 2016, <https://creativecommons.org/publicdomain/zero/1.0>.
7. Jackie Fahmy (knowledge base product analyst at OCLC) in discussion with the author, October 2015.
8. Ibid.
9. Liam Earney, "Maximizing the Knowledge Base: Knowledge Base+ and the Global Open Knowledgebase," *Insights* 26, no. 3 (2013): 244–49, <http://dx.doi.org/10.1629/2048-7754.71>.
10. Hans Lieder (head of the department for bibliographic services at the Berlin State Library) in discussion with the author, November 2015.
11. Evelinde Hutzler (librarian at Universitätsbibliothek Regensburg), e-mail message to author, November 19, 2015.
12. Benjamin Bober (librarian at ABES) in discussion with the author, December 2015.
13. Tomoki Ueno (librarian at the University of Electro-Communications) and Tomoko Kagawa (librarian at Ochanomizu University), e-mail message to author, November 16, 2015.
14. Bober discussion.
15. Ueno and Kagawa e-mail.
16. Yvette Diven (product manager lead for management solutions at ProQuest) in discussion with the author, October 2015.
17. Bober discussion.

# Knowledge Base Profiles

The current knowledge base market includes a wide range of proprietary and open-source products. Proprietary products usually support a wide range of library services, including management and discovery tools. These profiles attempt to capture a snapshot of the current knowledge base landscape and describe the functionality associated with products. Each supplier included responded to a short profile questionnaire provided by the author.

Some knowledge base providers were unavailable to complete the questionnaire, resulting in a few major products being omitted from the following list. These include the SFX Knowledge Base from Ex Libris; the Innovative Central Knowledge Base from Innovative Interfaces, Inc.; and JournalFinder from W. T. Cox.

## Commercial Knowledge Bases

### EBSCO Integrated Knowledge Base

#### ORGANIZATION NAME

EBSCO Information Services (<https://www.ebsco.com>)

#### ORGANIZATION DESCRIPTION

EBSCO is a leading producer and provider of content and services serving the needs of researchers from libraries of all types and sizes. EBSCO has developed an end-to-end open discovery services platform around EBSCO Discovery Service that supports all content types, features advanced search logic, delivers discovery and holdings management tools, and ensures extensibility with an array of third-party applications. The platform streamlines staff

functionality that directly impacts the user, with the EBSCO Integrated Knowledge Base playing a key role in supporting features such as holdings management, publication searching and browsing, OpenURL and direct linking to full-text, e-resource management, consolidation of COUNTER statistics, analysis, reporting, and in-workflow decision support.

#### NUMBER OF USERS

4,200 customers are using products that rely on the EBSCO Integrated Knowledge Base.

#### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

- OpenURL link resolver
- publication browse
- discovery service
- Google Scholar/PubMed holdings update services
- KBART-1 and KBART-2 holdings exports
- MARC records service
- usage consolidation
- e-resource management

#### DISTINCTIVE FEATURES

The EBSCO Integrated Knowledge Base is truly a global knowledge base representing over 10,000 databases and packages from over 1,400 providers. The integrated nature of the knowledge base with its identifier mappings allows EBSCO to automate holdings management for databases, e-journals, e-packages, and e-books ordered through EBSCO. Financial information, license terms, and access and registration information are also automatically updated for e-journals and e-packages ordered through EBSCO.

The same integration allows EBSCO to offer in-workflow decision support by providing access to COUNTER statistics, cost-per-use information, and analytics within the subscription workflow.

The link resolver and discovery service leverage an article-level knowledge base of over 120 million article links to offer a first-of-its-kind direct linking technology (introduced in 2001) that provides confirmed direct links to subscribed content, greatly improving the quality of linking and combatting a common problem of link resolvers where poor quality data on OpenURLs compromise link quality. This same technology allows EBSCOhost and EDS to provide access to more of the library's collection by integrating direct links to subscribed content into search results.

#### *FUTURE DEVELOPMENT PLANS*

EBSCO continues to focus development efforts on improving and expanding our knowledge base–related services. Future plans include

- improving librarians' user experience by offering a single interface for managing, reporting, and analyzing holdings, usage, licenses, and e-resource data
- expanding cost-per-use analysis and analytics to cover the entire collection
- supporting more COUNTER reports
- creating open integration with ILS systems to allow EBSCO and ILS partner systems to operate as one

### **Gold Rush**

#### *ORGANIZATION NAME*

Colorado Alliance of Research Libraries ([www.coalition.org](http://www.coalition.org))

#### *ORGANIZATION DESCRIPTION*

The Colorado Alliance of Research Libraries is a non-profit organization of fourteen research libraries in Colorado and Wyoming (thirteen academic and one public library) established in 1971 and incorporated as a non-profit 501c3 in 1981. The driving force is cooperation and the sharing of purchasing power, materials, and ideas. Among the services offered by the consortium are the Prospector union catalog, the Gold Rush ERMS, consortial e-resource licensing, a shared print program, and continuing education and training.

#### *NUMBER OF USERS*

About fifty libraries in North America use one or more modules of the Gold Rush ERMS.

#### *SERVICES SUPPORTED BY THE KNOWLEDGE BASE*

The Gold Rush service (<https://www.coalition.org/software/gold-rush>) includes a link resolver, A–Z service for serials, ERMS for managing subscriptions, and Gold Rush Decision Support. The service is centrally managed, and libraries may subscribe to any needed module at a cost far below commercial counterparts. The Gold Rush Decision Support supports a knowledge base of over 1,700 title lists, which include publishers, aggregators, abstracting and indexing services, and specialty lists (e.g., Portico, CLOCKSS, CrossRef, shared print serial sets, open-access lists, etc.).

#### *DISTINCTIVE FEATURES*

The Gold Rush Decision Support service allows libraries to do content overlap between electronic resource packages from primary publishers, aggregators, and indexing/abstracting services. Users can compare one-to-one or many-to-many in the same simple interface. Results are displayed in graphical form, and analyses can easily be downloaded as needed. Libraries may also load title lists from other services such as a commercial ERMS, RapidILL, or other sources that may be used for comparative purposes.

#### *FUTURE DEVELOPMENT PLANS*

A new area of development that was released in fall 2015 is the Gold Rush Library Content Comparison System (<https://www.coalition.org/faq-library-content-comparison-system>), which was developed to allow libraries to load their MARC records and compare them with other libraries in the system. It was initially developed to support the Shared Print program of the Colorado Alliance of Research Libraries but is now available to any library or consortium for a reasonable fee. There are many possible use cases for the system, some of which could include

- shared print programs among a group of libraries so that libraries can make better decisions about what to weed or put in storage
- adding a new program at an institution where the library wants to see how its collection compares to an institution that has a similar program in the same area
- a library loading a special collection of titles that are under consideration for weeding or storage to determine what is unique in that particular set
- performing quick exports of data sets for participation in other cooperative programs
- analyzing a collection for accreditation or membership in another organization

## ProQuest Knowledgebase

### ORGANIZATION NAME

ProQuest LLC (www.proquest.com)

### ORGANIZATION DESCRIPTION

ProQuest is committed to empowering researchers and librarians around the world. The company's portfolio of assets—including content, technologies, and deep expertise—drives better research outcomes for users and greater efficiency for the libraries and organizations that serve them. ProQuest is headquartered in Ann Arbor, Michigan, with offices around the world.

### NUMBER OF USERS

2,800+ libraries in more than 150 countries worldwide

### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

- 360 Core (A-to-Z list)
- 360 Link (link resolver)
- 360 MARC Updates (OPAC updating service)
- 360 Resource Manager (electronic resource management)
- Intota (library services platform)
- Intota Assessment (print and electronic collection analysis and assessment)
- Summon (discovery service)

### DISTINCTIVE FEATURES

At ProQuest, we have an integrated, centrally managed knowledgebase. From its origins in the year 2000 as “Serials Solutions KnowledgeWorks”—the first dedicated e-resource knowledgebase in the library industry—our knowledgebase has been a repository of high-quality, continuously updated metadata about e-journals, e-books, and other resources that is used across our services. For this reason, we can deliver consistent, synchronized metadata to any and all of the products that use the knowledgebase.

The fact that our knowledgebase is centrally curated and managed means that ProQuest libraries can utilize the same high-quality metadata across its librarian-facing tools (including ERM and assessment), as well as its discovery and access services. With our hosted software-as-a-service (SaaS) model, we make updates to the knowledgebase that are shared across all of our customers' services at once.

For the past fifteen years, ProQuest has developed and used increasingly comprehensive processes for cleaning, verifying, reconciling (“normalizing”), and updating the data we gather from content aggregators, hosts, publishers, and other providers. These

processes create a corrected and consistent set of metadata that can be used across our products so that librarians and researchers don't have to worry about the quirks or inconsistencies that are inherent in a surprising percentage of the source data.

### FUTURE DEVELOPMENT PLANS

Over the past three years, ProQuest has been hard at work behind the scenes, transforming our knowledgebase and expanding its scope, scale, and capabilities into a new, even more comprehensive knowledgebase. The new knowledgebase includes all of the e-resource metadata ProQuest curates, plus the serials and provider metadata we maintain in our Ulrich's Global Serials Directory, as well as our expansive store of MARC source records and data from new sources. The work we have accomplished enables us to bring together electronic, print, microform, and digital resource metadata in one place—on a new knowledgebase platform—and share it across a wider array of ProQuest services through APIs and web services. The new knowledgebase is also cloud-based, so we are able to innovate and scale the knowledgebase for future growth and expansion easily and effectively.

## TDNet Discover

### ORGANIZATION NAME

TDNet (www.tdnet.io)

### ORGANIZATION DESCRIPTION

TDNet is a leading provider of information technology solutions for libraries and knowledge centers. TDNet is dedicated to helping knowledge workers work faster and more efficiently while enhancing user experience. TDNet's highly flexible solutions meet the needs of individual libraries, knowledge centers, and consortia doing much of the work and saving both time and expenses. TDNet's company flagship—TDNet Discover—leverages years of experience and understanding of customer needs, reduces administrative workload, simplifies discovery, and enables library personnel to focus on serving their patrons.

### NUMBER OF USERS

Hundreds of customers worldwide

### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

- TDNet Discover—discovery web-scale search
- TDNet Discover—Library e-Resources—e-resources discovery and access gateway
- TDNet Discover—OpenURL link resolver

- TDNet Discover—TOC alerts service
- TDNet Core ERM—electronic resource management system
- TDNet Holdings Manager—MARC records and other knowledge base—extracted information service

#### *DISTINCTIVE FEATURES*

TDNet Discover uniquely combines technology and content, together with services. At TDNet, we believe that the search process and its results are a significant stage in a much broader and complex organizational process. Based on this approach, TDNet Discover is not a stand-alone platform but part of a collection of organizational research workflow tools and processes. As such, discovery-to-delivery must be adapted to the organization's entire work environment.

These are TDNet Discover's features that enable users to discover and access information in enterprise content repositories, external repositories, licensed and open-access publishers' content, the web, and more:

- full library portal with efficient information deployment
- advanced, comprehensive content and search capabilities
- multisite, consortia, group support
- extensive statistics reporting tools
- built-in SUSHI statistics harvester
- responsive interface for mobile
- compatibility with authentication protocols
- full interoperability with enterprise workflows and infrastructures and full API support

#### *FUTURE DEVELOPMENT PLANS*

Aiming to best serve our core customer base—corporate, biomedical, government, and other special libraries and information centers—TDNet's development road map follows the holistic approach of developing all components of our offering. We are pursuing continued development of our comprehensive knowledge base and index, optimization of search and retrieval processes and open-access exposure.

### **WorldCat Knowledge Base**

#### *ORGANIZATION NAME*

OCLC ([www.oclc.org](http://www.oclc.org))

#### *ORGANIZATION DESCRIPTION*

OCLC is a global library cooperative that provides shared technology services, original research, and community programs for its membership and the

library community at large. We are librarians, technologists, researchers, pioneers, leaders, and learners. With thousands of library members in more than 100 countries, we come together as OCLC to make information more accessible and more useful, because what is known must be shared.

#### *NUMBER OF USERS*

More than 4,700 total member libraries use the WorldCat Knowledge Base.

#### *SERVICES SUPPORTED BY THE KNOWLEDGE BASE*

As OCLC has built new services and transformed our foundational services for the age of electronic resources, the WorldCat Knowledge Base has been placed alongside WorldCat at the center of everything OCLC does:

- WorldCat Discovery (web-based discovery service)
- A-Z List (public-facing inventory of e-resources)
- WorldShare ILL (resource sharing service)
- WorldShare Acquisitions (ordering and procurement)
- WorldShare License Manager (license management and usage statistics solution)
- WorldShare Analytics (collection analysis tool)
- WorldShare Collection Manager
- MARC record delivery service

#### *DISTINCTIVE FEATURES*

The WorldCat Knowledge Base aggregates e-resource data from over 5,900 different vendors and provides link resolution for 3.7 million open-access titles. As a content-neutral knowledge base provider, OCLC is proud to work across the broadest possible range of vendors and content partners.

OCLC was first to implement direct holdings feeds from content providers into the WorldCat Knowledge Base, updating a library's coverage quickly and accurately, and that program continues to expand today. Partners in this program as of November 2015 include EBL Ebook Library, ebrary, Ingram MyiLibrary, Elsevier ScienceDirect (journals and e-books), JSTOR, and Teton Data Systems.

WorldCat Knowledge Base has been designed and deployed to be leveraged at any level the library needs and chooses. It can be integrated with OCLC applications like WorldCat Discovery or WMS, easily synchronized with another knowledge base, integrated with third-party applications as a data platform, or used to enrich data for use in external systems.

The WorldCat Knowledge Base is the first cooperatively managed knowledge base. Each institution has the option to deny or approve updates to collection data from vendors before they are loaded to



the knowledge base. Institutions can also contribute brand-new collection data, which the rest of community can then make use of. With the help of members OCLC is building a collaborative and comprehensive global knowledge base.

#### *FUTURE DEVELOPMENT PLANS*

With a goal of getting as close as possible to real-time updates, OCLC is continually investing in architecture and in exploration of better, faster methods of getting updates from partners. OCLC is committed to gaining new partnerships with vendors and implementing direct holdings feeds to create a “hands-off” e-resource management system for libraries. OCLC is also experimenting with an option to receive vendor data on demand through APIs instead of depending on file loading.

OCLC’s recent focus has been on improving the scalability of the system. OCLC is building a system to handle continual growth as the data ingested from providers and libraries continues to grow. Comprehensiveness is a goal libraries can achieve in cooperatively managing the WorldCat Knowledge Base.

The user experience is the ultimate goal of this work, and near-100 percent Google-style reliability of links is a critical component. Medium-term strategies include a move to direct linking to complement or in some cases supplant OpenURL linking. OCLC is currently testing a direct linking solution using Gale collection data and plans to expand this testing to other vendors.

## **Open Knowledge Bases**

### **BAsE de COnnaissance Nationale (BACON)**

#### *RESPONSIBLE ORGANIZATION*

Agence Bibliographique de l’Enseignement Supérieur (ABES; <http://en.abes.fr>)

#### *ORGANIZATION DESCRIPTION*

ABES was created in 1994 to implement Sudoc (Système Universitaire de Documentation, or University Documentation System), the union catalog of France’s higher education libraries. Sudoc opened in 2001 and has proved a resounding success. It covers the collections of 1,419 “deployed” or member libraries, along with the 1,793 public or private libraries from the Sudoc-PS network, which specializes in referencing serial publications. With over 10 million bibliographic records, 32 million localized documents and 24 million public queries in 2013, it plays a leading role in the French higher education and research information system.

#### *SERVICES SUPPORTED BY THE KNOWLEDGE BASE*

BACON provides trusted KBART v2 formatted metadata for e-resources packages available for French higher education institutions. These metadata, put under a CC0 license, can be downloaded via BACON’s website (<https://bacon.abes.fr>) and is accessible via web services. KBART files can be then used by knowledge base vendors and libraries.

#### *DISTINCTIVE FEATURES*

BACON focuses on French content. Data that can be fetched from other trusted community knowledge bases (KB+, GOKb) [is] integrated as is. For French content, we spend a lot of time encouraging French academic publishers to enhance their own metadata, and we insist that the KBART files be produced from the metadata used by the publishers’ platforms. We have built a semi-automated workflow that analyses the data sent to us by the publishers and converts it to trusted metadata sources (SUDOC, ISSN registry, French National Library catalog). We are then able to produce a detailed report that helps the publishers spot the mistakes or the inconsistencies of their metadata. If the publishers correct their metadata, ABES grants them a “quality label.” The major benefit for the publishers—and for everyone in the supply chain—is that the corrected and enhanced metadata can be used in any metadata feed, including ONIX files, MARC records, and data sent to discovery tools vendors.

#### *FUTURE DEVELOPMENT PLANS*

Future development plans include full automation of the file analysis workflow and full coverage of French academic publishers.

### **CUFTS Knowledgebase**

#### *RESPONSIBLE ORGANIZATION*

Simon Fraser University Library ([www.sfu.ca](http://www.sfu.ca))

#### *ORGANIZATION DESCRIPTION*

Simon Fraser University (SFU) is a medium-sized publicly funded institution serving a student population of approximately 19,990 FTE. SFU offers comprehensive undergraduate and graduate programs with three campuses located in the Metro Vancouver region of British Columbia, Canada. The SFU Library employs approximately 113 FTE personnel.

#### *NUMBER OF USERS*

Approximately 66

#### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

- GODOT: OpenURL link resolver and interlibrary loan-requesting software
- CJDB: CUFTS Journal Database, a public, web-based A–Z electronic journal listing
- integration with CUFTS ERM for public display of license information via the CJDB
- simple MARC record service (title, ISSN, e-ISSN, and holdings by provider on a single record) for import into integrated library systems
- import of print MARC journal holdings for integration into the CJDB A–Z public display
- automated monthly export of Google Scholar XML holdings for Google Scholar Library links
- automated monthly export of holdings for use in the BrowZine service
- CUFTS Resource Comparison Tool—compares up to four CUFTS targets in the knowledgebase to find duplicate and unique coverage
- Journal Search—finds out which CUFTS targets in the knowledgebase contain full text for a specific title
- off-campus authentication services (such as EZproxy or Innovative’s WAM) supported, and a proxy prefix can be added automatically by selecting proxy for each target

#### DISTINCTIVE FEATURES

Developed by an academic library for use in academic libraries in a consortia environment, the CUFTS knowledgebase is maintained by staff at the SFU Library. The knowledgebase contains the majority of the popular aggregator databases from EBSCO, Gale, and ProQuest as well as journal collections from large commercial academic publishers, university presses, and scholarly societies. In addition, the CUFTS open knowledgebase includes the Canadian Research Knowledge Network (CRKN) consortia journal packages. With Simon Fraser University Library’s commitment to establishing leading-edge scholarly communications support, significant efforts are made to populate the knowledgebase with open-access journal targets and free back issue targets. Open Journal Systems (OJS) targets are also well represented in the knowledgebase

All targets in the CUFTS knowledgebase display a “title list scanned” date, which provides the date the target was last updated. Whenever partially activated targets are updated in the global knowledgebase, the contact listed in CUFTS will receive an e-mail message detailing the number of new titles added, modified, and deleted during the update as well as tab-delimited text files for each of the new, modified, and deleted titles that affect the library’s holdings. Library contacts receive a deleted file only if any of their own activated titles were deleted by the global update.

The CJDB can also be integrated with the CUFTS Electronic Resources Management (ERM) module to display relevant license information for end users. License information appears in easy-to-read tabbed format and offers simple icons and plain language for end users and library staff. Some Canadian academic institutions have opted out of the Access Copyright agreement and rely on the Canadian Copyright Act and their own existing license agreements made directly with publishers and providers. So there has been an emphasis among Canadian academic institutions to make their license details publicly accessible.

#### FUTURE DEVELOPMENT PLANS

CUFTS is currently in a “steady state.” There is a committed user community, but it is not growing dramatically. Similarly, ongoing incremental development is always underway, but at present there are no plans for any major development initiatives.

#### Electronic Resources Database-JAPAN: ERDB-JP

##### ORGANIZATION NAME

A Working Group for E-Resource Data Sharing (<https://erdb-jp.nii.ac.jp/ja>)

##### ORGANIZATION DESCRIPTION

A Working Group for E-Resource Data Sharing was established by the Future Scholarly Information Systems Committee to handle ERDB-JP. The Future Scholarly Information Systems Committee operates under the Cooperation Promotion Council set up by the Inter-University Research Institute Corporation, the Research Organization of Information and Systems, the National Institution of Informatics (NII), and the Japanese Coordinating Committee for University Libraries.

#### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

Link resolver and web-scale discovery service

#### DISTINCTIVE FEATURES

ERDB-JP is a one-of-a-kind knowledge base describing electronic journals and books written in Japanese and electronic journals and books edited or published in Japan. ERDB-JP covers more than 11,000 journal titles as of October 2015.

#### FUTURE DEVELOPMENT PLANS

- *Quality improvement of ERDB-JP data:* We are continuing to evaluate the optimal maintenance

organization needed to provide accurate and current ERDB-JP data.

- *Increasing ERDB-JP partners:* ERDB-JP partners maintain ERDB-JP data along with the working group. We are encouraging electronic resources publishers, commercial knowledge base vendors, and academic conferences to consider ERDB-JP partnership.
- *International collaboration:* We are going to transmit ERDB-JP data to GOKb for the distribution of Japanese research outcomes.
- *Electronic books and licensing:* We are evaluating the possibility of adding collections of electronic books and electronic resources licenses to ERDB-JP.

## Global Open Knowledgebase (GOKb)

### RESPONSIBLE ORGANIZATIONS

The Kuali Foundation (<https://www.kuali.org>) and Jisc (<https://www.jisc.ac.uk>)

### ORGANIZATION DESCRIPTIONS

The Kuali Foundation is a nonprofit organization that develops open-source administration software for higher education. Kuali is also the parent organization to Kuali OLE, a community source library management system and sister project to GOKb. Jisc is a not-for-profit organization that supports digital services and solutions for the UK higher education sector. Jisc Collections supports the Knowledge Base Plus (KB+) project, also a project partner to GOKb. Kuali OLE and Jisc Collections have been working together since 2012 to develop GOKb as an open, community-managed knowledge base to support the broader community as well as their own individual projects.

### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

GOKb aims to make knowledge base data freely available to the library community and provide the infrastructure necessary for partners to participate in the data management process. While GOKb does not support typical knowledge base-powered tools such as a discovery platform or ERMS, its open data and APIs are designed to allow external systems to consume the data in support of these functions.

Features include

- a web interface for browsing and searching data
- editor functionality that allows GOKb partners to deposit new data, correct errors, and contribute data enhancements like title history information
- OAI-PMH standards-based APIs designed for easy consumption and integration of data

### DISTINCTIVE FEATURES

In addition to traditional knowledge base metadata, GOKb offers an enhanced data model that tracks changes over time, relationships between resources, and an extensible set of external identifiers. A co-referencing service within the knowledge base allows users to submit an identifier and receive a results set of all known identifiers associated with the same resource, through either the web interface or an API. All of the data in GOKb can be accessed through the web interface, API, or export tools.

The data found in GOKb is completely managed by the GOKb partners, which include the Kuali OLE partners, Jisc, and a number of additional library partners with an interest in the service. GOKb's data is openly available under a CC0 license. It can be used by anyone, for any purpose, without attribution. Academic institutions and commercial publishers and vendors are encouraged to collaborate in building and sharing GOKb's data.

### FUTURE DEVELOPMENT

The GOKb development team is completing several development initiatives as part of its second round of grant funding from the Andrew W. Mellon Foundation. Features planned for release in 2016 include support for e-book packages, more advanced data-loading and management tools, and exposure of the knowledge base as linked data. GOKb will also continue to engage in community-building activities and is actively seeking new partnerships with libraries, consortia, publishers, and vendors.

## Knowledge Base Plus (KB+)

### RESPONSIBLE ORGANIZATION

Jisc (<https://www.jisc.ac.uk>)

### ORGANIZATION DESCRIPTION

Jisc is the UK higher, further education, and skills sectors' not-for-profit organization for digital services and solutions.

### SERVICES SUPPORTED BY THE KNOWLEDGE BASE

KB+ is a knowledge base that includes electronic resources management tools. All of the KB+ data is made available under an open license and disseminated throughout the library supply chain so that the right organizations have the data they need when they need it. Currently Ex Libris, ProQuest, OCLC, and EBSCO all use KB+ data in their systems. KB+ data is also used by other Jisc services or projects including JUSP and Safenet.

#### *DISTINCTIVE FEATURES*

- A centrally maintained and managed knowledge base in which Jisc Collections collates, verifies, and updates knowledge base data to avoid costly and wasteful duplication of effort by libraries all trying to do the same thing by themselves.
- Verified, accurate, and up-to-date publication information for e-journal agreements, including national and regional consortium agreements from across the United Kingdom and a growing number of non-Jisc packages.
- Subscription information and management tools to help institutions track details of entitlements and journal coverage, manage renewals, compare

different journal packages, view usage statistics from JUSP, and export files formatted for use with link resolvers.

- License information covering key values such as walk-in users, concurrent access, post cancellation access, and more. Institutions can create their own license information, making use of templates created by Jisc Collections or their own licenses.

#### *FUTURE DEVELOPMENT PLANS*

Incorporation of financial data will enable measurement of value (i.e., cost per use) and assessment of the strategic value of a title on a dimension other than raw usage.

# Library Technology

R E P O R T S

## Upcoming Issues

October 52:7	<b>Google Analytics and Google Tag Manager</b> by Tabatha Farney
November/ December 52:8	<b>Social Media Optimization: Improving the Visibility and Use of Library Resources</b> by Scott W.H. Young and Doralyn Rossmann
January 53:1	<b>Internet of Things: Mobile Technology and Location Services in Libraries</b> by Jim Hahn

## Subscribe

[alatechsource.org/subscribe](http://alatechsource.org/subscribe)

## Purchase single copies in the ALA Store

[alastore.ala.org](http://alastore.ala.org)



[alatechsource.org](http://alatechsource.org)

ALA TechSource, a unit of the publishing department of the American Library Association