

# 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,

Published From	1975/01/01		
Published To	1988/12/31		
Title History	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
	1988-01-01	• Physica B+C	• Physica B: Condensed Matter • Physica C: Superconductivity

**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.

Tetrahedron	Title Name	Search	Show: Selected and Not Selected	Limit To: All
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		
AGORA		FAO (Food and Agriculture Organization)		
<a href="#">CIBER-CASPUR</a>	01/01/1957 - Present	CASPUR		
<a href="#">Cilea 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>.