Components of E-resource Access

The technologies employed to deliver library e-resource access to end users have evolved considerably over the past decade. Within the academic library sphere, online public access catalogs (OPACs) and federated search interfaces have given way to "web-scale" index-based discovery systems; e-resource holdings and linking information are now administered within cloud-hosted knowledge management systems rather than locally hosted integrated library systems (ILSs); and user authentication has expanded to include a variety of IP and federated identity management (FIM) options. In order to facilitate the discussion around e-resource access disruptions, we begin with a chapter on these technological developments. In this chapter, we define the technology components through which library end users gain access to electronic materials, focusing on those that comprise the discovery service environment. We describe how each component works, the role it plays within the larger library system, and how metadata from these key systems plays an integral role in e-resource access. We also discuss the different types of metadata, the systems from which they originate, and the spheres of control that govern their management.

Search and Discovery

Library systems consist of four basic components: search and discovery (access) tools, knowledge management systems, linking systems, and authentication. Regardless of how a library configures its system, these four pieces must be present to enable e-resource access. We begin by discussing search and discovery.

Terminology

- *Access tools,* sometimes called discovery or retrieval tools, are any computer application through which a library user can discover and gain access to an e-resource. Features and functionality vary greatly from tool to tool, and libraries typically employ multiple tools to meet a variety of access needs. Types of access tools include online public access catalogs, database A–Z lists, e-journal A–Z lists, and web-scale discovery services.
- *Central or discovery indexes* are collections of "pre-harvested and processed metadata and full text that comprises the searchable content of a [web-scale discovery] service" (Hoeppner 2012, 7). These indexes harvest and normalize vendor-supplied resource data, which can include "rich" metadata such as abstracts, author-supplied keywords, tables of contents, and full text. It is these indexes that power web-scale discovery services.
- *Database A–Z lists* are alphabetical lists of databases (and other selected e-resources) to which a library provides access. Libraries create these lists through a variety of methods, which range from manually adding hyperlinks to a static web page to developing a homegrown database solution to employing a vendor product, such as Springshare's LibGuides A–Z Database List.
- *Discovery interfaces* are search applications that ingest and index metadata from a variety of sources, including institutional repositories, digital collections, and APIs (Breeding 2018). They provide users with advanced search features, such as keyword recommenders, limiters, facets, and relevancy ranking of results. These features are meant to encourage more serendipitous discovery rather than strict known-item retrieval.
- *Discovery services,* sometimes called index-based discovery services or web-scale discovery services, are products that combine a discovery interface with a central index. Unlike a standalone discovery interface, a discovery service facilitates the discovery of resources outside of a library's holdings via its connection to a central

or discovery index or indexes. It also allows for article-level and chapter-level search results and linking.

- *E-journal A–Z lists* are alphabetical lists of electronic journals to which a library provides access. These lists are typically auto-populated according to the library's holdings. Besides acting as a searchable inventory of a library's e-journals, an A–Z list also collates and displays each e-journal's available access points, as well as other relevant information, such as coverage dates and notes regarding licensing and authentication.

Discussion

Index-based discovery services have become the most widely adopted discovery application by academic libraries. Previously, most libraries employed online public access catalogs through which library users could search locally maintained metadata records. OPACs were quickly found to be insufficient to support e-resource access because these resources morphed and multiplied more rapidly than individual libraries could maintain them. This created constant errors and inaccuracies within OPACs and led to frustration by librarians and library users alike. Discovery services, by contrast, reduce the pressure on individual libraries to keep up with the constant flux of e-resource metadata. By utilizing repositories of e-resource metadata compiled and maintained by a discovery service vendor, libraries are able to provide more robust and up-to-date coverage of their e-resource holdings, as well as delivering a more granular (and Google-like) search experience to users.

The discovery service market is dominated by a handful of commercial vendors that host and maintain the discovery service on behalf of their library customers. Discovery service search results are populated from centralized indexes, which have ingested and normalized data from hundreds of publishers, aggregators, and content providers. Content included in these indexes comes from both open-access and commercial sources and encompasses everything from e-books and e-journals to video, images, sound recordings, government documents, and more. Discovery services also facilitate the discovery of local catalog and institutional repository records, which can be contributed by the library via FTP or OAI-PMH protocol.

Because central indexes harvest metadata from hundreds of content providers, many of which have their own standards for representing e-resource information, the accuracy and quality of the ingested metadata vary from provider to provider. Similarly, what and how much data is shared by content providers is governed by their contracts with the discovery service vendor. Some content providers, for instance, authorize their data to be utilized only by subscribing institutions. Discovery service vendors that also act as content providers (e.g., EBSCO and ProQuest/Ex Libris) are unwilling to exchange metadata in order to preserve a competitive edge for their discovery product. This has led to opaqueness around both the discoverability of e-resources within a library's chosen discovery service and how the robustness (or meagerness) of the data within the central indexes has influenced e-resource usage.

Academic libraries have supplemented their use of discovery systems with additional access tools for more targeted discovery needs. OPACs, for instance, are sometimes employed in tandem with a discovery service and are used primarily for known-item searching. Other common access tools used by libraries include database A–Z lists, which are popular for giving end users an easy-to-scan list of their library's available online databases. E-journal A–Z lists fulfill a similar function for the discovery of electronic journals, allowing for the easy search and retrieval of known serials titles. These access tools are maintained either independently by the library (as with Springshare's LibGuides A–Z Database List) or as part of a broader knowledge management system, which we discuss next.

Knowledge Management Systems and Link Administration

Terminology

- *Direct linking* refers to the creation of links within a discovery service by leveraging provider-specific, proprietary metadata from the central index. Direct linking is usually employed by discovery services alongside link resolver/OpenURL linking because it "provide[s] more reliable access to electronic resources than through the OpenURL process" (Breeding 2018, 7).
- An *ERMS*, or electronic resource management system, is a knowledge management system that specializes in tracking and managing electronic resources throughout their life cycle. An ERMS is typically powered by a centralized knowledge base, which allows librarians to easily find and activate specific instances of e-resources or e-packages, and includes additional management features, such as the ability to store payment, licensing, and contact information; to receive renewal reminders; and to track usage.
- An *integrated library system* is a suite of modules used by librarians to manage the activities involved in acquiring and loaning materials, such as ordering, invoicing, cataloging, and circulation. ILSs were originally developed to provide

operational support for physical materials and as a result are poorly equipped to handle the complexities of e-resource management. These inadequacies have prompted the development of other tools, such as ERMSs, A–Z lists, and library services platforms (LSPs).
- *Knowledge bases* are centralized databases of metadata that describe specific instances of e-resources available through a publisher, content provider, or platform (Wilson 2016). A knowledge base includes not just basic bibliographic information (title, author, publisher, etc.) but also information about the resource's platform, vendor, coverage dates, and access model, including which packages or collections it appears in. Knowledge bases are used to power a variety of knowledge management systems and access tools. The primary purpose of the knowledge base is holdings management, allowing libraries to track which e-resources they have with certain vendors. This, in turn, supports the article-level links users encounter in a library's discovery service and the title-level links in a library's A–Z lists.
- *Link resolver,* or OpenURL linking, refers to the "specialized software used to provide context-sensitive links among the panoply of systems that compose a modern library's electronic collections" (Chisare et al. 2017, 93). Utilizing the OpenURL encoding format, link resolvers create their links by combining the citation data of the desired resource (source) from a library discovery record with the provider website (target) linking parameters necessary to connect to the desired resource. For a link resolver to know which resources are locally available to a library user, it must be connected to a knowledge base that has been pre-populated with a library's electronic holdings.
- *LSP* refers to a next-generation library system that incorporates the functionalities of an ILS, a knowledge base, a link resolver, and an ERMS. Library services platforms were developed as a way to unite the disparate knowledge management systems into one comprehensive system and support the workflows of electronic, digital, and physical material.

Discussion

As e-resources increased in availability, it quickly became clear that integrated library systems were inadequate to support the maintenance of electronic holdings. While e-resource MARC records could be loaded into ILSs, the accuracy of these records decreased as the overall number of records increased. Vendor participation in holdings workflows was often limited to supplying a library with MARC records, and these records frequently needed remediation to bring them up to cataloging standards. Thus, the onus of holdings maintenance rested entirely on local libraries. The sheer volume of data that needed to be maintained quickly became overwhelming for libraries without the staff or time available to offset the cumbersome workflows.

The proliferation of electronic resource management systems in the mid-2000s further enticed libraries away from traditional models of holdings management. ERMSs are stand-alone systems connected to a link resolver knowledge base, which provided context-sensitive links to e-resource content. The advent of link resolvers and their attached knowledge bases became a panacea for the historical efforts of loading individual MARC records for e-resources. Companies such as Serials Solutions provided knowledge bases that could be used to track the collections, packages, and individual subscriptions available to a library. These knowledge bases also could be connected to a discovery service to provide a single-search experience for users to find both e-resource and print content, as well as retrieve more granular results, such as at the article or chapter level.

While a mix-and-match approach to discovery is available, libraries tend to procure their ERMS, link resolver, and discovery service as a suite of products from the same vendor. This trend of bundling services is likely to continue into the foreseeable future as the discovery industry continues to consolidate, leaving libraries with fewer vendors to choose between. Next-generation library systems take this one step further with the library services platform, which combines the functionality of an ERMS/knowledge base with that of a traditional ILS, providing a unified place to administer both print and electronic resources. While LSPs are still in their infancy, they promise to reduce the number of disparate systems needed by electronic resources librarians to effectively manage their e-resources.

Authentication

Terminology

- *Authentication* is the process of proving one's identity as an authorized or legitimate user of a product or service. Most vendors and content providers require that users first prove their affiliation with the purchasing or subscribing library before they are allowed to access content on the platform. Libraries employ various methods of authentication, including via IP address, proxy server, virtual private network (VPN), and single sign-on (SSO).

- *Proxies* are a type of intermediary server or software system that sits between one computer and another. Libraries commonly employ proxies to authenticate remotely located users because a proxy enables a library to override a computer's IP address with its own, thus changing the computer's apparent location. The most commonly employed proxy system for libraries is EZproxy.
- *Federated identity management*, or federated SSO, refers to a system of single sign-on that enables users to authenticate into applications across multiple unrelated third-party domains using a single set of credentials. With federated identity management, a user's credentials are verified by a trusted identity provider (often the user's educational institution), which then communicates the user's authentication status to third parties via a secure protocol, such as SAML or OAuth. FIM enables library users to authenticate into multiple content provider platforms using a single set of credentials without the need for IP addresses, proxies, or VPNs. Common identity federations include InCommon (for Shibboleth SSO) and OpenAthens.
- *Multifactor authentication*, or two-step authentication, is an authentication method in which a user verifies their identity using additional pieces of information beyond their username and password. This information may be the answer to a security question, a security code sent to a verified e-mail address or phone number, or acknowledgment of the log-in attempt via a third-party application.
- *Single sign-on* is a form of authentication that uses session information stored as a cookie on a web browser to automatically authenticate a user into multiple applications within the same organization after the user has logged in once. Single sign-on is frequently used by higher education institutions to reduce the number of times a user needs to authenticate into applications hosted or provided by the institution. It is increasingly used in conjunction with multifactor authentication to provide added account security.
- *VPNs,* or virtual private networks, are services that create a secure, encrypted connection from one computer to another. Similar to a proxy, a VPN acts as a middleman for a computer and its destination, sitting between them and overriding the connecting computer's IP address with its own. However, unlike a proxy, a VPN is more secure because it encrypts a computer's information before it even connects to the internet.

Discussion

IP authentication is currently the most popular way to authenticate library users. When a library acquires an e-resource, it provides the vendor with a set of IP ranges that represent the library's computer and Wi-Fi network. When a user connects to the e-resource over the internet, the vendor checks the device's IP address to see if it falls within the provided ranges. If it does, the user is granted access. If not, the user is redirected to an error or a payment message. Since this process happens behind the scenes, the user is never prompted to enter credentials, making the movement from discovery record to e-resource appear seamless. Unfortunately, IP authentication by itself is able to provide access only for users who are currently located on the library's or institution's physical site. As a result, IP authentication is frequently used in conjunction with other authentication methods to grant access to users who are located remotely.

Many libraries employ a proxy service jointly with IP authentication to enable e-resource access to users located outside the library's physical premises. When a remotely located user attempts to connect to an e-resource through one of the library's access tools, the browser is redirected to the proxy server, which asks for the user's credentials. The browser redirect can happen a couple of different ways but typically involves modifying the e-resource's URL, such as adding a prefix to the beginning of the e-resource's URL. Once the proxy verifies the user's credentials against its internal database, it connects the browser to the desired resource using its own IP address. Since the proxy server's IP address is included in the authorized ranges given to vendors, the user is granted access to the e-resource. In addition to the proxy prefix, a proxy requires maintenance of several configuration files to function, including one that contains the URLs, hosts, and domains of the e-resource's platform. The configuration file needs to be frequently updated to keep pace with vendor platform developments.

Another way to provide access to remote users is through a VPN, or virtual private network. A VPN fills a similar role as a proxy, acting as an intermediary between the user's device and the desired e-resource. Just as with a proxy, a user's device must first connect with the VPN, thus assuming its IP, before connecting to the e-resource. Because the VPN's IP address is included in the ranges provided to the vendor, the device appears to be located on site and is authorized for access. However, unlike a proxy, a VPN requires users to download and install specialized software onto their personal devices, configuring it with settings specific to their institution. But not all institutions' VPN services are configured to provide access to e-resources. Some institutions implement a practice called split tunneling, which means the VPN routes only certain types of web traffic through its server, while the rest access the internet normally. Institutions that use split tunneling generally route only traffic destined for internal resources, such as those

Table 2.1. Sources and types of metadata

Component	Sources of Metadata	Types of Metadata	Sphere of Control
Online Catalog/ILS	• Original MARC record cataloging • Individual MARC record loads • Bulk MARC record loads	• Bibliographic metadata • Database/collection citation metadata • Book citation metadata • Journal citation metadata • Video citation metadata • URLs	Library
Central Index	• Data supplied by publishers, vendors, and content providers	• Bibliographic metadata • Video citation metadata • Article citation metadata • Abstracts • Full text • Direct links • DOIs • Table of contents	Vendor
Knowledge Base	• Data supplied by publishers, vendors, and content providers	• Bibliographic metadata • Database/collection citation metadata • Book citation metadata • Journal citation metadata • Video citation metadata • Parser & parser parameters • Link resolver information	Vendor
Discovery Service	• Online catalog/ILS • Central index • Knowledge base • APIs	• Bibliographic metadata • Database/collection citation metadata • Book citation metadata • Journal citation metadata • Video citation metadata • Article citation metadata • Abstracts • Full text • Direct links • DOIs	Blended
Library Services Platform	• Original MARC record cataloging • Individual MARC record loads • Bulk MARC record loads • Knowledge base	• Bibliographic metadata • Database/collection citation metadata • Book citation metadata • Journal citation metadata • Video citation metadata • Parser & parser parameters • Link resolver information • Site IDs	Blended
Link resolver	• Knowledge base	• Citation information • Parser & parser parameters • Link resolver information	Vendor
ERMS	• Selection of holdings from a knowledge base	• Bibliographic metadata • Database/collection citation metadata • Book citation metadata • Journal citation metadata • Video citation metadata • Site IDs	Blended
Database A-Z List	• Manual record creation	• Database/collection title • URLs	Library
E-journal A-Z List	• Auto-populated from holdings selected from a knowledge base	• Journal citation metadata • Holdings/coverage dates • URLs	Blended
Research Guide	• Manual entry • Asset management tool	• Database/collection title • Book title • Journal title • Video title • URLs	Library

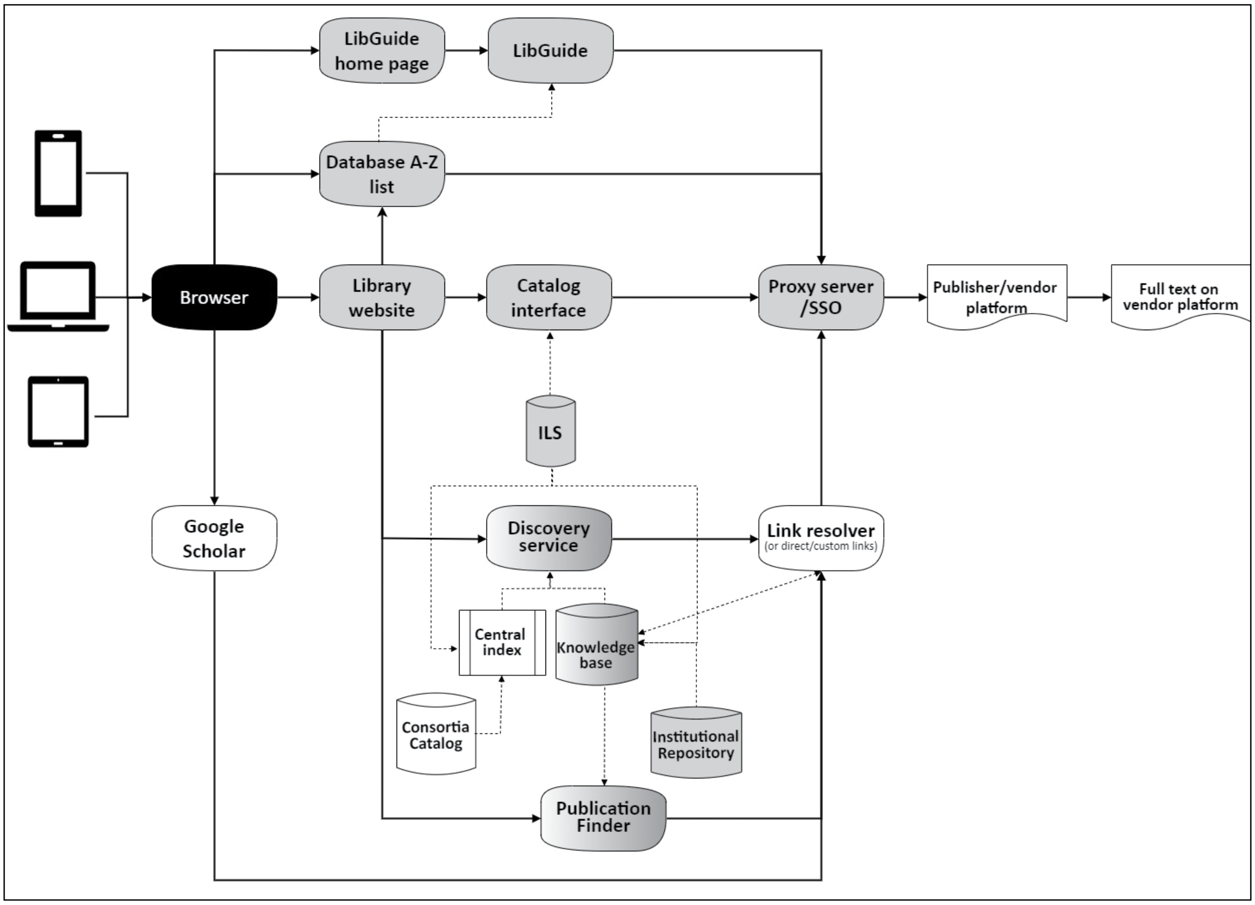

Figure 2.1
Comprehensive access chain. Black = patron-controlled metadata, system, or tool; gray = library-controlled metadata, system, or tool; white = vendor-controlled metadata, system, or tool; gradient indicates shared control.

hosted on the institutions' intranet, through the VPN; all other traffic, including that going to library e-resources, accesses the internet using the user's normal router and IP address. This means if the user is off site, they will not be authenticated correctly.

While IP authentication remains widely employed by academic libraries, federated identity management (FIM) authentication continues to grow as a preferred method of authentication by academic libraries and vendors due largely to its ability to provide more account security, such as through multifactor authentication. With FIM authentication, a user can navigate to an e-resource from anywhere on the internet, including Google, and be able to log in by choosing their institution from the provided drop-down menu, often called a WAYF (Where Are You From), on the vendor's platform. Once a user logs in, the information (called a token) is stored as a cookie on the browser, which can then be shared by other resources and vendors without the user needing to log in again. Because FIM requires vendors to join an identity federation, such as InCommon (Shibboleth) or OpenAthens, as well as install and configure additional software on their servers, not every vendor will have it as an option. As a result, FIM is often used in conjunction with other authentication methods such as proxy to provide robust coverage.

Sources and Types of Metadata

A significant portion of e-resource access disruptions is derived from incorrect metadata. Bibliographic, holdings, and platform information form the backbone of all library access and linking tools. This means any missing, erroneous, or out-of-date metadata will adversely affect the discoverability of an e-resource and potentially lead to breakdowns in access. However, metadata can originate from a number of sources, including internally within the library or externally with a publisher, content provider, or discovery vendor. It is also often blended together within individual access tools, making it difficult to pinpoint where the metadata came from, what portion is causing an access issue, and which party is responsible for correcting it. Understanding the flow of metadata from its various origination points is therefore essential.

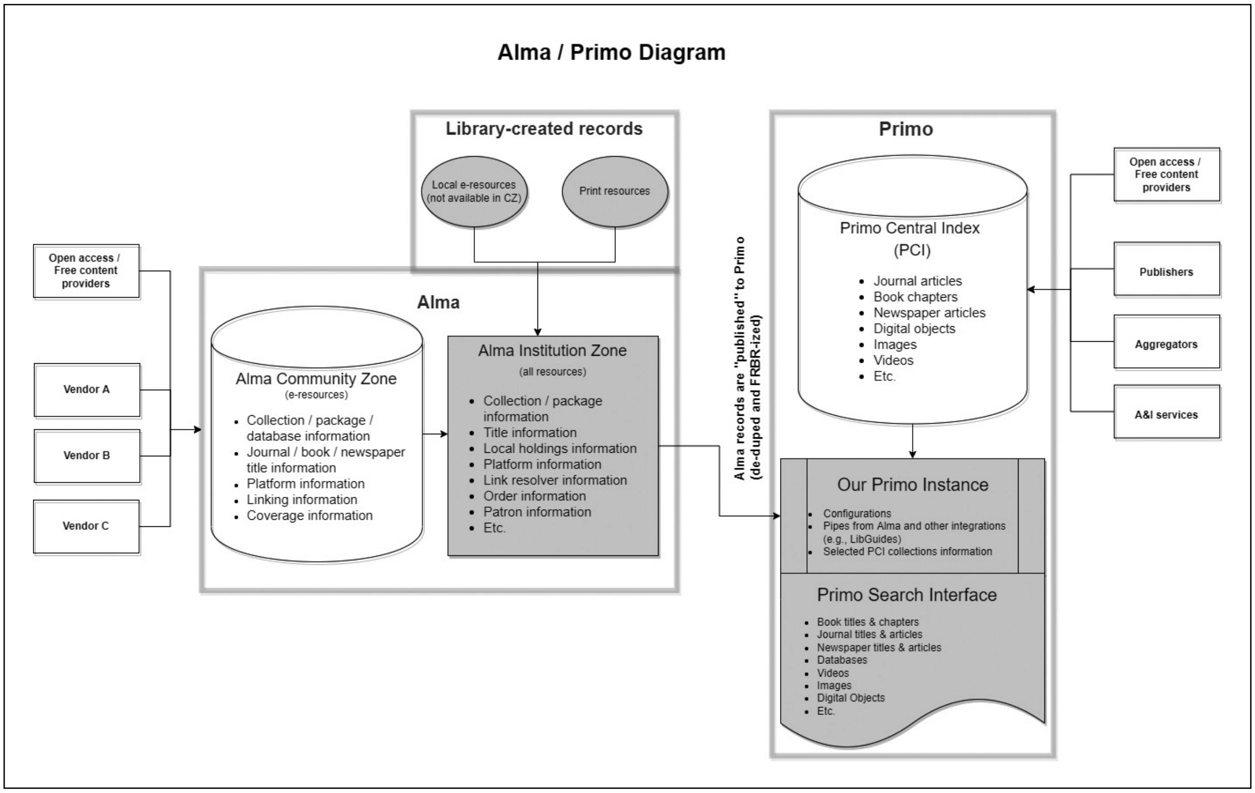

Figure 2.2
Library services platform: Alma/Primo access chain. White = Ex Libris or vendor-controlled metadata; gray = library or blended controlled metadata.

Table 2.1 (p. 10) summarizes the sources and types of metadata that feed into each component in the comprehensive access chain. We have also included a rough guide to whose sphere of control each falls under: library, vendor, or a blend of the two. This distinction is important because depending on whose sphere of control the component falls under, a troubleshooter will have a greater or lesser ability to test hypotheses, effect change, and enact solutions. This table is solely focused on e-resource metadata and therefore does not take into account other sources of print, digital, or institutional repository metadata. Also, please note that the table is not exhaustive and represents only metadata found to be the most commonly used for diagnosing e-resource access disruptions.

Comprehensive Access Chain

Figure 2.1 (p. 11) depicts how search and discovery (access) tools, knowledge management systems, linking options, and authentication methods work together to enable access to a library's electronic resources. The diagram details a few paths a user may take through the chain of access (solid line), as well as the flow of metadata between the various components (dotted line). It includes an example of how users can begin their discovery journey outside of the library website with Google Scholar, which can be configured to utilize the library's link resolver to connect users to the library's holdings. Other abstract and indexing (A&I) databases offer similar functionality, but it is up to individual subscribing libraries to decide which platforms it is enabled on. Figure 2.2 depicts how the same technology components are utilized in a library services platform, in this case Ex Libris's Alma/Primo.

References

Breeding, Marshall. 2018. "Index-Based Discovery Services: Current Market Positions and Trends." *Library Technology Reports* 54, no. 8 (November/December): 1–33. https://doi.org/10.5860/ltr.54n8.

Chisare, Cyndy, Jody Condit Fagan, David Gaines, and Michael Trocchia. 2017. "Selecting Link Resolver and Knowledge Base Software: Implications of Interoperability." *Journal of Electronic Resources Librarianship* 29, no. 2: 93–106. https://doi.org/10.1080/1941126X.2017.1304765.

Hoeppner, Athena. 2012. “The Ins and Outs of Evaluating Web-Scale Discovery Services.” *Computers in Libraries* 32, no. 3 (June 24): 6–40. https://www.infotoday.com/cilmag/apr12/Hoeppner-Web-Scale-Discovery-Services.shtml.

Wilson, Kristen. 2016. “The Knowledge Base at the Center of the Universe.” *Library Technology Reports* 52, no. 6 (August/September): 1–35. https://doi.org/10.5860/ltr.52n6.