

# Introduction

Libraries strive to implement the best technologies to provide access to the collections they build on behalf of their sponsoring institutions or communities. These tools must be convenient for users yet deliver sophisticated features and return relevant and objective results. The realm of library discovery products has become increasingly multifaceted as it evolves to meet the expectations of librarians and library users. The information environment these discovery tools address has become increasingly complex and reshaped by multiple models of open access publishing and competitive dynamics among the major companies involved in publishing, scholarly workflows, and analytics.

This issue of *Library Technology Reports* gives an updated look at the realm of discovery products implemented in libraries, focusing especially on how these products have been implemented in academic libraries. It is the third issue of *Library Technology Reports* produced by this author addressing topics related to library catalogs or discovery services.

“Next-Generation Library Catalogs” (July/August 2007) detailed the transition from traditional online catalogs provided with integrated library systems (ILSs) to a new genre of discovery interfaces designed to accommodate the expectations of users acclimated to the more elegant and powerful interfaces of popular internet services.<sup>1</sup> This work is now dated but remains as a historical overview of the emergence and development of the first generation of discovery interfaces apart from library catalogs.

“Library Resource Discovery Products: Context, Library Perspectives, and Vendor Positions” (January 2014) provided an overview of the many types of discovery products available at that time and included survey results from libraries evaluating their

perceptions of the effectiveness and objectivity of these products.<sup>2</sup>

This issue focuses primarily on index-based discovery services. This genre of products was established in 2009 and has since become a mainstay of academic libraries. Despite broad interest, the number of players in this product category has remained limited and constant. Four products were launched in 2009: Summon from Serials Solutions/ProQuest, WorldCat Local from OCLC, Primo from Ex Libris, and EBSCO Discovery Service from EBSCO Information Services. No new products have entered the field. The merger of Ex Libris into ProQuest has brought about a consolidated process for producing their respective indexes, but Summon and Primo Central continue to be offered as ongoing products.

Almost a decade has transpired since the introduction of these products. Libraries have made a substantial economic investment during that period. These products have gained general acceptance in academic libraries as one of the components expected among their service offerings, often the launchpad offered to their users to gain access to their full array of content and service options.

The immense investment in these products warrants a look at some of the patterns in which they have been implemented in libraries, which may in turn inform trends to expect looking forward. In the absence of comprehensive and reliable data regarding the deployments of these products globally, the author gathered data describing the use of these products among colleges and universities in the United States. This group of libraries forms an important constituency for these products. Although patterns may vary within each global region and among other types of libraries, these US academic libraries can be taken as generally

representative of the broader market.

This issue of *Library Technology Reports* gives a high-level view of the general characteristics of the index-based discovery services and the overall marketplace trends. It does not aim to provide a detailed evaluative look at the features provided by each product. The marketplace study included in chapter 4 can help libraries understand which products have been most successful among a library's peer institutions.

## Terminology and Definitions

Before we turn to the discussion of the products and technologies, the following definitions will clarify some terminology that is often used inconsistently in vendor discussions and in the professional literature. These definitions apply specifically to the library context and may be used differently in other disciplines or types of organizations.

**Discovery**, in the library context, describes the general activity of locating library resources, primarily by its users or members. In current times, discovery tends to be accomplished primarily through search technologies, but it can also include browsing or other serendipitous means of presenting resources to users based on their interests. Searching can take place through an interface or tool offered by the library as well as through services provided by other entities.

**Discoverability** involves techniques that enable library resources to be more optimally found in the search or discovery environments of organizations outside the library. A library resource management system, for example, can publish instances of an item or its metadata encoded in a linked data schema optimized for harvesting by general or scholarly search engines. Discoverability is an example of search engine optimization applied to library resources.

**Delivery**, or **access**, describes the mechanisms or processes to enable items found through an act of discovery to be made available to the user. The appropriate method for delivery to a user depends on the format of the item and any applicable copyright or license arrangements that may limit access. Access to electronic resources may be mediated through a link resolver that directs the user to a version of the item available within the library's subscriptions or through alternate services. Materials restricted to subscribers may require that the user be validated through a personal or institutional authentication service. Libraries strive to make the delivery of materials to which their users are entitled optimally transparent, with any needed authentication and linking taking place behind the scenes without intervention. Delivery mechanisms for physical materials involve established processes for lending or interlibrary loan.

**Discovery interface** describes applications created to enable users to discover and gain access to library resources. Typical features of a discovery interface include core search and retrieval features for accessing library resources, query recommendations through type-head or drop-down selections, relevancy ranking of results, presentation of facets to narrow search results, and other interface tools and conventions to facilitate the search process. Discovery interfaces do not come with prepopulated indexes of content but rather ingest content from external sources such as the organization's ILS, local institutional repositories, and digital collections, or from external services through APIs (application programming interfaces). Discovery interfaces usually operate with ILSs and other repositories provided by different vendors or developers. Discovery interfaces usually include internal indexing technologies, such as Apache Solr or Elasticsearch. Examples of discovery interfaces include

- Encore from Innovative Interfaces, which operates with a Millennium or Sierra ILS and local repositories. A version of the product branded as Encore Duet enables libraries subscribing to EBSCO Discovery Service to incorporate article-level results.
- Enterprise from SirsiDynix, which can be used with the company's Horizon or Symphony ILS product. SirsiDynix has a partnership with EBSCO Information Services to include article results for mutual customers.
- VuFind, an open source discovery interface originally developed at the Villanova University Falvey Memorial Library.
- Blacklight, an open source discovery interface originally created at the University of Virginia and Stanford University.

**Index-based discovery services**, sometimes called web-scale discovery services, are products that include a prepopulated central index along with a specialized discovery interface. These discovery services enable libraries to provide article-level results to their users. The major index-based discovery services include

- EBSCO Discovery Service from EBSCO Information Services
- Primo from Ex Libris
- Summon from Ex Libris
- WorldCat Discovery Service and its predecessor WorldCat Local from OCLC

These products depend on the major publishers, aggregators, producers of abstracting and indexing (A&I) services, and other providers to provide

resources that can be added to the central index. A key issue for these discovery services relates to the currency of the index. Publishers provide new resources to discovery services periodically, but not necessarily as the resources are added to their own servers. This workflow of content means that the central index of a discovery service may have delays in providing access to recent articles.

An index-based discovery service returns descriptive records with elements such as citation data, abstracts, and tables of contents. Once a record is selected, users click a link or button to access the full text of articles from the server of the publisher or aggregator. This workflow enables publishers to maintain control of their content as users discover resources through library-provided discovery services.

**Central index**, in the context of index-based discovery service, is a large-scale index, created and maintained by the discovery service vendor. Vendors of discovery services make arrangements to receive content from publishers and aggregators to build these indexes, which ideally represent the body of resources of interest to libraries that are made available through institutional subscriptions or open access licenses.

**Federated search**, synonymous with **meta-search**, describes a method for enabling users to search multiple content resources through a single search query. These products used protocols such as Z39.50 to transmit the user's query to multiple search targets and to present the results to the user. Results from the targets might be interleaved with each other, according to factors such as publication date or relevancy, or grouped by targets. These products were popular in the library arena from about 2000 until 2009. Metasearch products, though a pragmatic solution, were not especially well regarded due to slow performance, delivery of shallow result sets limited by the time and bandwidth required to transfer all the records returned by a query for each target, and non-intuitive interfaces. This genre of discovery product faded following the launch of index-based discovery services. Since federated search applications receive results directly from the servers of the information provider, even the most recently added resources are retrieved, unlike index-based discovery, which depends on periodically ingested materials.

A **link resolver** participates in the discovery service through facilitating access to resources returned in search results or from citations in documents. These products rely on the OpenURL standard for creating metadata-enriched links, a knowledge base of e-resource holdings, a profile of the library's subscriptions, and business logic to parse the link and, based on its metadata, connect the user to the version of the resource that the library subscribes to or that is available as open access. Link resolvers emerged to address the need to create reliable linking to articles that can

be managed at large scale rather than individually. Updates to the link syntax on a publisher site, changes in subscriptions, or other scenarios that might otherwise require massive manual intervention can be accomplished through configuration changes in the link resolver.

**Smart linking** is often implemented in discovery services to provide more reliable access to electronic resources than through the OpenURL process, making use of internal or proprietary data beyond what would be available through OpenURL.

**APIs**, or **application programming interfaces**, provide access to functionality and data from one computer system to another. APIs consist of a specific set of commands or directives, formulated according to specified syntax and structure, that enable a computer application to return a response for a request submitted through a program script or other application. While user interfaces enable humans to operate a software application, APIs work behind the scenes. Most discovery services offer APIs that enable a third-party discovery interface to initiate a search query and receive search results based on its central index and relevancy algorithms. In addition to search results, a discovery environment may also use APIs from an ILS or library services platform for features related to patron profiles, circulation, and personalization.

A **knowledge base**, in the context of library discovery, manages details related to packages of library resources. Content managed within a knowledge base generally includes the individual journal titles and coverage dates for each resource package offered to libraries and any additional relevant details, such as title or publisher changes. The commercial knowledge bases attempt to describe all of the content packages of interest to libraries, including subscription-based and open access. Knowledge bases can be filtered by a profile of a library's subscriptions and entitlements to determine the availability of an article. OpenURL link resolvers make use of a knowledge base as the key component in the process of linking to full text, which usually depends on whether the desired content item is covered within the library's body of subscriptions.

**Integrated library systems** (ILSs) provide operational support to libraries, including modules for cataloging library materials, managing acquisitions, circulating materials, and enabling users to search and make requests through an online catalog. The model of automation seen in ILSs was established in the late 1970s. The ILS emerged at a time when libraries were involved in print collections and has not been adapted to managing electronic resources. These products continue to be dominant in public and school libraries.

**Library services platforms** embrace a more comprehensive model for the management of library collections, designed to accommodate the workflows of electronic, digital, and print materials. These

products are deployed on web-based, multi-tenant platforms and provide internal knowledge bases and other shared content components. This genre of products, launched in about 2011, includes Alma from Ex Libris and WorldShare Management Services from OCLC. The FOLIO project is underway to produce an open source library services platform, with implementations expected by early 2019. Unsuccessful efforts to develop other library services platforms include the Kuali OLE open source project and Intota from ProQuest. The need for ProQuest to complete Intota was obviated when it acquired Ex Libris and its already well-established Alma product. Library services platforms have been adopted mostly by academic and research libraries, though the conceptual design could also be applied to other types of libraries.

## Special Challenges in Addressing Open Access Content

The default model of discovery provides access to resources within the library's body of subscriptions. Link resolvers are well positioned to be able to determine if an article should be available within subscribed resources. All persons are also entitled to the full text of any article published as open access, regardless of whether their library has purchased a subscription to the journal. Discovery services and link resolvers need some type of indicator that an article may be available as open access. Some of the techniques used to identify eligible open access content include the "Access and License Indicators" recommended practice from NISO (RP-22-2015), which defines metadata elements reflecting the status of an article.<sup>3</sup> These elements can be used in many ways, including submissions to discovery indexes and knowledge bases to better enable access to open access materials.

Another method for identifying open access materials involves the use of specialized citation databases that track these resources. The nonprofit organization Impactstory, for example, has created Unpaywall, a browser plug-in that helps researchers identify any available PDF full-text copies of resources they search for on the web.<sup>4</sup> Unpaywall captures the citation and checks it against the library's subscriptions as well as its database of open access resources available on institutional and disciplinary repositories. Link resolvers can also be configured to enable the Unpaywall service to provide links to open access copies of articles that may be outside the library's subscriptions.

Link resolvers from Ex Libris, including SFX, 360 Link, and Alma's resolver, support access to the Unpaywall service.<sup>5</sup> EBSCO offers an app that can be installed in a library's instance of EBSCOhost or EBSCO Discovery Service to check the availability of open access articles for items listed in search results.

*EBSCO Unpaywall API*

<https://cloud.ebsco.com/apps/unpaywall-api>

## Models of Discovery

Libraries today have many different options to enable their users to discover and gain access to their collections of information resources. Several different combinations of products can be assembled, depending on the ways in which the library wants to organize its website and discovery environment and the types of resource management systems it has in place. The following section describes some of the combinations currently seen on academic library websites in the United States.

### Online Catalog with No Index-Based Discovery

Libraries with an online catalog with no index-based discovery present a search box for the online catalog for access to locally owned items such as books, DVDs, and other materials managed by their ILS. Although most midsized and large academic libraries have implemented one of the commercial index-based discovery services, many smaller institutions direct users to specific aggregated databases or lists of e-journals. Public libraries predominantly present the online catalog of the ILS as their main search box.

### Index-Based Discovery Service with Separate Online Catalog

Another set of libraries have implemented an index-based discovery service but use it primarily for access to their electronic resources while maintaining the online catalog of their ILS for books and other local resources. Libraries often use a tabbed search box on their website to direct a query to the appropriate system. This approach makes it simple for a user to enter search terms, but once the initial search has been submitted, the user is then working with the native interface of the online catalog, discovery service, or other system.

### Bento Box

The bento box search model involves a discovery interface designed to offer a single box for users to enter a query, which is then simultaneously sent to multiple systems. Results are then organized into multiple panels, each representing the items returned from one service. The term *bento box* was coined by Tito Sierra, using the metaphor of a bento box keeping different types of food items separate to visualize the organization of different types of information separated into

cells of a webpage.<sup>6</sup> A typical bento box would present results from the library catalog, articles from the discovery service, content from the website, and resources from a digital collection within separate panels. The bento box discovery interface would interact with the APIs of each of the target services to populate each panel of the results page. Users can then select any panel to see a more complete list of resources according to the content type of interest.

### Online Catalog Integrated with Index-Based Discovery

Academic libraries operating an ILS may want to continue to use their online catalog as their discovery interface and be able to include articles in search results. This search model is accomplished through the use of the API of the discovery service called within the online catalog or discovery interface of the ILS. Almost all major ILS developers, for example, have created a mechanism for integrating with EBSCO Discovery Service (EDS). Encore Duet from Innovative Interfaces is based on Encore as the discovery interface, retrieving results from Sierra or Millennium for locally owned items and layering in articles through the EDS API. SirsiDynix follows a similar approach with its Enterprise discovery interface.

### Discovery Service with ILS Integration

The model of discovery service with ILS integration presents the interface of the discovery service to users, integrating results from its central index along with results from the library's ILS. The discovery interface can serve as a complete replacement for the online catalog of the ILS, including features related to the patron account such as placing requests for materials, viewing lists of items borrowed, and other related features.

### Bundled Discovery Service with Library Services Platform

Bundled discovery services with library services platforms provide a unified model for managing library resources across multiple formats. These products are used by library personnel to support the operations

of the library and for managing collection resources. They differ substantively from ILSs, which provide resource management through modules and workflows originally designed for print resources. The two major library services platforms, Ex Libris Alma and OCLC's WorldShare Management Services, do not directly include a patron-facing catalog or discovery service. Both organizations, however, bundle their discovery products with their library services platform. Although library services platforms can also be implemented with alternative discovery services, most installations to date pair components from the same vendor.

### Fully Integrated Discovery Service with Library Service Platform

Another model of discovery, a fully integrated discovery service with library service platform, brings the discovery service closer to its associated library services platform. Even when bundled together, Primo and Alma have been managed through separate configuration interfaces (or "back office" tools, to use Ex Libris terminology). Ex Libris recently launched a new option with a tighter coupling of Primo and Alma, both managed through the Alma back office.

### Notes

1. Marshall Breeding, "Next-Generation Library Catalogs," *Library Technology Reports* 43, no. 4 (July/August 2007).
2. Marshall Breeding, "Library Resource Discovery Products: Context, Library Perspectives, and Vendor Positions," *Library Technology Reports* 50, no. 1 (January 2014).
3. "NISO RP-22-2015 Access and License Indicators," Recommended Practice, NISO, January 5, 2015, <https://www.niso.org/publications/niso-rp-22-2015-access-and-license-indicators>.
4. Holly Else, "How Unpaywall Is Transforming Open Science," *Nature* 560 (August 2018): 290–91 <https://doi.org/10.1038/d41586-018-05968-3>.
5. See Heather, "oaDOI Integrated into the SFX Link Resolver," *Impactstory Blog*, February 2, 2017, <http://blog.impactstory.org/oaDOI-in-sfx>.
6. Cory Lown, Tito Sierra, and Josh Boyer, "How Users Search the Library from a Single Search Box," *College and Research Libraries* 74, no. 3 (2013): 227–41.