



Primo Search

CONFIDENTIAL INFORMATION

The information herein is the property of Ex Libris Ltd. or its affiliates and any misuse or abuse will result in economic loss. DO NOT COPY UNLESS YOU HAVE BEEN GIVEN SPECIFIC WRITTEN AUTHORIZATION FROM EX LIBRIS LTD.

This document is provided for limited and restricted purposes in accordance with a binding contract with Ex Libris Ltd. or an affiliate. The information herein includes trade secrets and is confidential.

DISCLAIMER

The information in this document will be subject to periodic change and updating. Please confirm that you have the most current documentation. There are no warranties of any kind, express or implied, provided in this documentation, other than those expressly agreed upon in the applicable Ex Libris contract. This information is provided AS IS. Unless otherwise agreed, Ex Libris shall not be liable for any damages for use of this document, including, without limitation, consequential, punitive, indirect or direct damages.

Any references in this document to third-party material (including third-party Web sites) are provided for convenience only and do not in any manner serve as an endorsement of that third-party material or those Web sites. The third-party materials are not part of the materials for this Ex Libris product and Ex Libris has no liability for such materials.

TRADEMARKS

"Ex Libris," the Ex Libris bridge, Primo, Aleph, Alephino, Voyager, SFX, MetaLib, Verde, DigiTool, Preservation, URM, Voyager, ENCompass, Endeavor eZConnect, WebVoyage, Citation Server, LinkFinder and LinkFinder Plus, and other marks are trademarks or registered trademarks of Ex Libris Ltd. or its affiliates.

The absence of a name or logo in this list does not constitute a waiver of any and all intellectual property rights that Ex Libris Ltd. or its affiliates have established in any of its products, features, or service names or logos.

Trademarks of various third-party products, which may include the following, are referenced in this documentation. Ex Libris does not claim any rights in these trademarks. Use of these marks does not imply endorsement by Ex Libris of these third-party products, or endorsement by these third parties of Ex Libris products.

Oracle is a registered trademark of Oracle Corporation.

UNIX is a registered trademark in the United States and other countries, licensed exclusively through X/Open Company Ltd.

Microsoft, the Microsoft logo, MS, MS-DOS, Microsoft PowerPoint, Visual Basic, Visual C++, Win32,

Microsoft Windows, the Windows logo, Microsoft Notepad, Microsoft Windows Explorer, Microsoft Internet Explorer, and Windows NT are registered trademarks and ActiveX is a trademark of the Microsoft Corporation in the United States and/or other countries.

Unicode and the Unicode logo are registered trademarks of Unicode, Inc.

Google is a registered trademark of Google, Inc.

Copyright Ex Libris Limited, 2009. All rights reserved.

Document released: Month Year

Web address: <http://www.exlibrisgroup.com>

Table of Contents

1	Primo Search	4
	Native Primo Search	7
	<i>Local Primo Index</i>	7
	<i>Remote Primo Index</i>	8
	<i>Primo Central</i>	8
	Primo Deep Search	8
	Primo Metasearch	9
	Primo Link	10
	Primo—the Best of All Worlds	11

Primo Search

The information resources that libraries make available to users are varied. In their effort to provide a comprehensive selection of quality information resources, libraries offer their own collections—physical or digital—as well as collections from sources such as other libraries, primary and secondary publishers and aggregators, organizations holding repositories of scanned e-books, and Web sites.

Librarians face numerous challenges in choosing appropriate information resources, licensing them when necessary, and presenting these heterogeneous collections to patrons in a compelling and easy-to-use way. For most libraries, a single point of access to all of their collections is the ultimate goal.

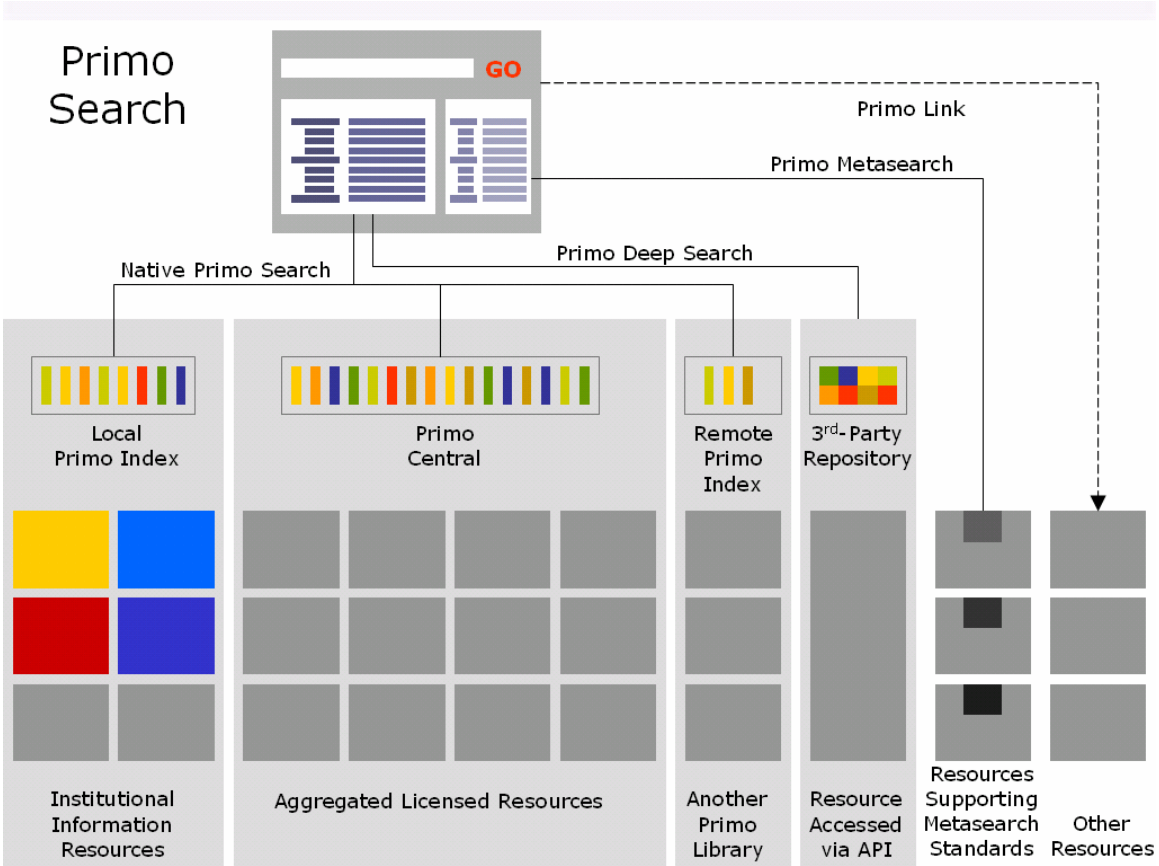
The Ex Libris Primo® discovery and delivery solution meets this goal by providing several complementary technologies that leverage a library's unique collections and position these collections prominently in the information landscape of the library. A library can deploy any combination of the Primo technologies; the ones that it chooses depend on the specific collections that it wants to offer its users, the technical resources that are at its disposal, the collaborative relationships that it can establish with other institutions, and its agreements with information providers. Regardless of the technologies deployed, users have a similar, though not identical, experience when looking for information across all resources.

In deploying Primo with its range of search technologies, libraries have full control over the selection and description of their collections, the way in which the collections are presented, and the discovery and delivery options offered to users. Based on an infrastructure that is independent of commercial publishing and handles the full spectrum of scholarly materials, Primo encourages collaboration among libraries, information providers, vendors, and other stakeholders. With Primo, libraries benefit from other collections without having to replicate the work done elsewhere and can focus their energy on exposing their unique collections, which they can contribute to the community effort to provide as much scholarly data as possible to users.

The search technologies that Primo enables libraries to deploy are as follows:

- “Native” Primo search, enabling Primo to search in one or more Primo indexes and blend the results into one result list. Several deployment options are available for the “native” Primo search technology, and an institution can choose to deploy one or more of them:
 - Search in a local Primo index, an index that is derived from the library catalog and other collections under the governance of the library or the institution
 - Search in a remote Primo index, an index of another Primo institution
 - Search in Primo Central, a centralized, hosted Primo index that encompasses data harvested from primary and secondary publishers and aggregators
- Primo Deep Search, which enables Primo to perform optimized just-in-time searches in remote resources that are not indexed by Primo and are compliant with the Deep Search requirements
- Primo Metasearch, which enables Primo to perform just-in-time searches in remote resources that are not available through Primo Central or through Primo Deep Search. The Primo Metasearch technology uses standard metasearch protocols.
- Primo Link, which enables Primo to direct users to the native user interface of a specific information resource

This document briefly describes the various Primo search technologies and explains how only by offering a central index and a combination of such search technologies can a search solution respond completely to the needs of academic and research institutions that rely on a multitude of information resources, information providers, material types, and access methods.



Native Primo Search

The Primo publishing platform creates and maintains a Primo index by harvesting and processing bibliographic metadata and any other available data stored in local or remote resources. Primo can divide its index into several nodes, or “sub” indexes, to improve load balancing and to distribute data-harvesting responsibility among various collections. Because Primo can perform distributed searches in multiple Primo indexes, it can search simultaneously in very large institutional collections, in collections that other institutions control, and in a central repository managed by Ex Libris or a third party. Primo blends the results originating from multiple Primo indexes and displays them as one result list.

Local Primo Index

By harvesting and processing metadata from a library’s catalog, its digital collections, its course materials, and any other harvestable information resources, Primo creates a local Primo index for the library. The preprocessing of the bibliographic data and the complete control that the library has over the hardware and software that it uses to create and maintain the local index enable the library to present its collections to best advantage and to optimize the searching capabilities in these collections for the user community.

In the process of building a Primo index, a library can apply various configuration and adjustment options to normalize, augment, deduplicate, and group the harvested bibliographic records. Furthermore, the library can control the information that is used for the search mechanism, for the display, and for the faceted categorization of result lists, and it can tailor the relevance ranking of the results to the library’s profile.

The Primo search engine is based on the Lucene search routines, which have been enhanced to take full advantage of the specific nature of library materials. The indexing and the search process take into account metadata standards, such as MARC, Dublin Core, and MAB; library-oriented classification, such as the various subject headings; and relationships between items, such as various expressions of a single work. As a result, Primo offers sophisticated searching and an extremely rapid response time.

To create a local Primo index, the library needs to harvest, process, and store metadata. The library might also choose to harvest the primary data—for example, from local repositories of articles. However, because institutions offer a wide range of collections, harvesting all of them to bring them into the local Primo index is not always feasible. Ex Libris optimizes the harvesting of information resources that are relevant for all Primo customers by offering a centralized index, Primo Central.

Remote Primo Index

The Primo solution can use Primo indexes of other institutions—remote Primo indexes—as well. By configuring Primo to aggregate results from multiple Primo indexes and blend them into one result list, institutions can extend their offering and leverage the unique collections of other institutions, thus broadening the scope of scholarly data in a simple and cost-effective manner.

Primo Central

Primo Central is a centralized Primo index that is hosted and managed by Ex Libris and offered to Primo institutions.¹ The information that is harvested and described in the Primo Central index—primarily bibliographic metadata, abstracts, and the full text of scholarly materials—is of great relevance to the global Primo community.

Technologically, managing Primo Central is similar to managing any other Primo index at an institution; however, the Primo Central index is managed by Ex Libris in collaboration with information providers—primary and secondary publishers and aggregators—for the initial and ongoing harvesting and processing of their data and for the handling of authorized access to licensed resources. Primo uses the harvested data to provide the discovery capabilities to users; however, delivery of the full text is carried out by the information provider's system. Where applicable, the delivery is facilitated by an institutional link resolver such as SFX.

The Primo Central index is readily available to all Primo customers and is highly scalable, covering millions of scholarly materials, primarily articles and e-books. When required, authentication and authorization apply when a user accesses this central repository to obtain information based on the subscriptions of the user's institution.

The Primo search engine uses the Primo Central index along with the local Primo index. The search results from the central repository are blended with the results from the local index to form one set of results. Primo optimizes searches in both the local index and the centralized index and provides performance that is similar to that of searching only in a local index.

Primo Deep Search

With the Primo Deep Search technology, an institution can offer access to remote, third-party information resources that are not available for Primo Central indexing but are Deep Search compliant. Using this technology, searches in remote resources resemble searches in the

¹ Note: Primo Central is currently under development.

institution's local Primo index: the response time is very short; the result lists are sorted by relevance, enabling Primo to combine them with results obtained from the local Primo index, the Primo Central index, and other remote Primo indexes; and relevant facet information, when provided by the remote resources, is available to Primo for display.

Libraries can deploy the Deep Search technology to access information resources that comply with the following requirements:

- The response time is less than a second.
- The results are sorted by relevance.
- The information resource's index can be accessed directly through an application programming interface (API).
- (Strongly recommended) Facet information is provided, including facet titles and the number of records per title.

The Deep Search component provides an efficient way to access remote resources that Primo does not harvest for Primo Central, such as the DTV Article Database Service (DADS) repository of tens of millions of scholarly articles, maintained by the Technical University of Denmark and accessed by Primo at the Royal Library of Denmark; and the repository of 1.8 million e-books at Duxiu.com, accessed by Primo at Shanghai Jiao Tong University. The deployment of the Deep Search technology requires a library to obtain the cooperation of information providers in furnishing the requisite API.

Primo Metasearch

To make access possible to remote resources that a library cannot or does not want to harvest and are not available through Primo Central or Primo Deep Search, the Primo solution includes a metasearch component—the metasearch engine of the Ex Libris MetaLib® gateway and metasearch system. This component enables a patron to use the Primo search interface to launch a search in multiple heterogeneous resources. The metasearch component, employing standard protocols or special programs, converts the unified Primo query to the diverse search syntaxes that are supported by the information resources in which the user wants to search; delivers the query to the search engine of each search target; and collects the results once they are available at the search target. Primo then processes the results and either combines them with the results obtained from the Primo indexes or displays them as a separate result list.

The metasearch technology has been deployed by Ex Libris since 2001 and relies on a comprehensive knowledge base that enables the MetaLib metasearch engine to invoke a search in hundreds of databases. Implemented by 1500 institutions around the world, MetaLib is one of the most widespread metasearch systems in use today.

Primo Link

Primo Link enables users to discover information resources in Primo, link to a resource, and search in the resource's native interface.

Achieving a single access point for searching in all the information resources that a library offers its users is not always possible, given the number and diversity of resources. Even when used together, the local and remote Primo indexes, Primo Central, Primo Deep Search, and Primo Metasearch might fall short in addressing the technological, legal, or other aspects of some resources—for example, subject guides and resources that do not have a bibliographic structure, such as data sets. Furthermore, users sometimes want to access an information resource that addresses a specific need or that provides unique services. Primo, as the library's information gateway, grants such direct access. By harvesting the institution's SFX and MetaLib knowledge bases, Primo obtains a list of institutional e-resources (e-journals, e-books, and databases) and incorporates them in the Primo index, thus enabling users to look for information resources and link directly to them.

Primo—the Best of All Worlds

No single technology can cover all the discovery needs of an institution. Primo offers a spectrum of technologies and deployment options, as well as a centralized index, that complement each other and together provide a compelling and friendly search environment.

A combination of all these options provides a complete environment and enables libraries to emphasize their uniqueness and to maintain their control over a number of factors: the selection of resources; the way in which resources are presented and made available to users; and the institution's investment in collections, hardware, software, and the maintenance of systems and data, be it a local or hosted environment.