

Key Issue

Discovery services: next generation of searching scholarly information



CHRIS KEENE

Technical Development Manager
University of Sussex Library

Introduction

The new breed of resource discovery services is an evolutionary step forward in providing library users with a 'one-stop shop' where they can find information sources for their research. They provide a single search box that can search a library's online and physical content including articles, books, journals, newspaper articles, e-books, specialist collections and more. These discovery services have built on the concepts of both federated searching and next-generation catalogues (described briefly below).

Federated search services allow a user to enter a search query which is submitted to a number of external indexing databases, publishing platforms and academic resources. The results are aggregated from the various sources to provide one unified set. Federated search became a popular solution within academic libraries over the last decade as it negated the need for users to separately search many different databases and interfaces, making the job of finding research material quicker and easier. However, federated searching was hampered by different databases being inconsistent in the results they returned for a given search term, while each database would return full details for only a limited number of items, making it impossible for the system to provide true relevance ranking or refinement options. The process of connecting to a number of external databases was also often slow and unreliable.

Next-generation catalogues – the other development from which discovery services evolved – was a term coined to describe a new breed of online public catalogue interface¹. The web catalogues bundled with library management systems (LMS)

had generally fallen behind the more dynamic developments of the web at large. They were often characterized by being slow, with a dated interface and lacking the features commonly found on similar web services. A new breed of web interface started to emerge, addressing many of these shortcomings. These included those developed by LMS providers as additional products, such as Enterprise from SirsiDynix, and those from third parties independent of the underlying LMS, such as Aquabrowser and the open source VuFind.

A discovery service builds on both these services to provide a single, Google-like search box, which searches across a vast knowledge base of scholarly content as well as across local holdings, most notably the library catalogue. The results are returned from one underlying index, making it quick to display, additionally providing full facets for refinement such as item type, subject area and date. Items in the results can be flagged to show if the full text is available, either through subscription, purchase or free access.

Users are able to perform a search using simple keywords (as they – especially students – have come to expect on the web) and discover a broad range of materials for their research, without needing to learn complex search syntax or know the most relevant databases and journals to refer to. These systems have adopted practices commonly found on popular websites including alternative search suggestions, auto-completion of search queries, RSS feeds and customized mobile interfaces. Discovery services commonly provide functionality to save lists, export to popular bibliographic formats and customize to fit needs, though the

extent varies between particular products. Overall, the end experience is a huge leap forward from what was previously offered by federated search technology.

Knowledge is King

The core of a discovery service is the underlying proprietary knowledge base, holding the metadata for the articles, books, documents and other materials that the service can search across. The breadth and depth of the knowledge base is the killer feature that these services possess. Their success depends on it, and suppliers compete to hold the largest number of quality records.

To build a knowledge base, vendors have developed relationships with content producers and content platforms to obtain item metadata information for articles, books, theses, newspaper articles or other content. De-duplicating the merged index that is the result of this process is an important element. Suppliers have taken differing approaches to this, for example, one merges information for the same item into one master record, while another selects the highest quality record to search and display. Indexes are not always limited to metadata, with some systems able to search across the full text of some items.

Libraries have expressed concern that researchers need to know which sources are included or not when they are performing a search. Some system suppliers have been reluctant to provide information as to which publishers and platforms they have agreements with, and which journals are included. Reasons include licence agreement confidentiality and concern that such information will be used by competitors. However, such moves do not help relieve concerns over coverage. Traditional specialist abstracting and indexing services still have a clear role to play, especially for those with more advanced research needs.

The knowledge bases of these products are complemented with local holdings, including the library catalogue, special collections and institutional repository. This merging of local and online content to provide one set of search results should not be overlooked and is a positive innovation in what can be offered to users when looking for scholarly information.

The players

There are four main players currently in the market: Primo from Ex Libris, Summon from Serials Solutions, EBSCO Discovery Service and WorldCat Local from OCLC. The distinction between 'discovery service' and 'next-generation catalogue' can be grey and some products could be classed as either.

Primo from Ex Libris² was released in 2007 as a next-generation catalogue with the capability to index data from a range of sources. Libraries using the software were able to import data from external sources (such as publishers) but often used mainly to index local records such as the library catalogue and repository, and instead using the federated search functionality to allow users to search across e-resources. In 2010, Ex Libris launched Primo Central, a hosted knowledge base of articles and other scholarly content. The discovery service is provided by the combination of Primo, the software and local indexes, and Primo Central, the large knowledge base of content. Primo is the only offering mentioned here that can be hosted locally by the institution, though it, like the others mentioned here, can also be hosted by the supplier.

Summon³ was launched by Serials Solutions in the summer of 2009. Being the first product of its kind, it created much talk within the community. Summon is a hosted service that searches its own knowledge base (500 million items) and local content including the library catalogue. Summon is designed in such a way that the back end, the search engine that performs the search and returns data, is separate to the front end, providing the user interface. An institution can therefore choose to disregard the supplied user interface and create their own, using the service's API (application programming interface: allows one system to exchange information with another) to link to the back end.

EBSCO Discovery Service (EDS)⁴ was released at the end of 2009. It boasts similar features to those above and a fully integrated federated search component to complement the system's knowledge base. The product offers many options to customize the look and branding to meet local needs, and can search within the full text of some articles.

Worldcat Local⁵ was launched in 2007. It builds on OCLC's Worldcat service but, as the name suggests, is localized to show local holdings first in search results. While Worldcat is well known for its large source of book bibliographic data, Worldcat Local includes over 482 million journal articles.

What's in a name?

There are currently a number of names for what is essentially the same type of service. Names include 'vertical search', 'web-scale search' and 'resource discovery service'. Web scale originated from the technology world to describe services where the very notion that they could be affected by peak demand or increasing users/data was unthinkable⁶. Amazon and Google are good examples of this. However, the term has been adopted by the library community, led by OCLC, who use it as an umbrella term for the architecture of a number of their services⁷. *Library Technology Reports* from the ALA recently dedicated an issue to 'Web Scale Discovery Services'⁸. Vertical search is a term used to describe services which search a defined set of content or specific subject domain, as opposed to broad search engines such as Google⁹. Resource discovery service, or simply discovery service¹⁰ is the third common name being applied to these services, and the one used here.

The future

This is a young and rapidly developing marketplace. While some products have evolved over a number of years, the discovery service features mentioned here have only emerged in the last couple of years. The software-as-a-service (SaaS) approach of most of these products allows for constant development and change which, while generally welcomed by customers, can also lead to frustration if not carefully planned and communicated. Suppliers are announcing new customers on an almost weekly basis and knowledge base item counts continue to rise rapidly.

Discovery services are not without their issues, though many of these vary between products. Performing a search for a known item such as *Pride and Prejudice* can sometimes return a vast number of results, yet the book by Jane Austen (in English) may, frustratingly, not be anywhere near the top of

the results list¹¹. Likewise, the large number of articles in the index can make searching for journal titles a challenge. There is a worry that journals, databases and items which are not included in the knowledge base will be invisible to users. Finally, licencing metadata, as noted above, is key to a functional service. While suppliers have all made good progress in negotiating agreements, with lines between software vendors and content providers now less distinct than they used to be, and further mergers always a possibility, conflicts of interest can emerge. How stable will these agreements be in the future?

Discovery services are still relatively new products on the market, yet already they have seen a high level of interest and take-up by libraries – something we can expect to continue in the future. What will the future hold? It will, of course, be down to the suppliers of these products, but we can expect to see mobile technologies playing an increasingly important role. Integration with the local LMS to provide full borrower account functionality (e.g. renewals and reservations) would allow libraries to use the discovery service as a full catalogue replacement, avoiding the need for users to use two different systems. We can also expect the expansion of the accompanying knowledge base to be a continuing priority.

Discovery services have changed the discovery landscape and the relationship between software vendors, publishers and libraries. As such, they are now an important element within the online information community.

References and further reading

1. Breeding, M, *Next-gen library catalogs*, 2011, London, Facet Publishing.
2. Ex Libris – Primo:
<http://www.exlibrisgroup.com/category/Primo>
Overview (accessed 6 June 2011).
3. Serials Solutions – Summon:
<http://www.serialssolutions.com/summon/>
(accessed 1 June 2011).
4. EBSCO Discovery Service:
<http://www.ebscohost.com/discovery/eds-about>
(accessed 1 June 2011).
5. OCLC – WorldCat Local:
<http://www.oclc.org/worldcatlocal/default.htm>
(accessed 1 June 2011).

6. Lorcan Dempsey's Weblog: Web scale
<http://orweblog.oclc.org/archives/001238.html>
(accessed 21 June 2011).
7. Web-scale Management Services:
<http://www.oclc.org/webscale/> (accessed 1 June 2011).
8. Vaughan, J, Web Scale Discovery What and Why,
Web Scale Discovery Services, 2011, 47 (1), pp 5–11.
9. InfoWorld: What is Vertical Search
<http://www.infoworld.com/t/platforms/what-vertical-search-057> (accessed 1 June 2011).
10. Federated Search Blog: Discovering discovery services
<http://federatedsearchblog.com/2009/07/19/discovering-discovery-services/> (accessed 6 June 2011).
11. Stainthorp, P, How commercial next-generation library discovery tools have nearly got it right:
<http://paulstainthorp.com/2011/05/17/how-commercial-next-generation-library-discovery-tools-have-nearly-got-it-right/> (accessed 1 June 2011).

Further reading

Unified Resource Discovery Comparison: articles on discovery
<http://pinboard.in/u:chriskeene/t:resourcediscovery/>

Key Issue © Chris Keene