

## **Discovery layers and discovery services**

CHRISTISON, Andrew

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/7435/>

---

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

### **Published version**

CHRISTISON, Andrew (2013). Discovery layers and discovery services. Catalogue and Index, 170, 2-12.

---

### **Copyright and re-use policy**

See <http://shura.shu.ac.uk/information.html>

Sheffield Hallam University is currently undertaking a review of next-generation library management systems (LMSs). As part of that exercise I have considered discovery services and interfaces on the market and looked at developments in discovery/OPAC display. This article reviews the main options available and offers suggestions as to what factors libraries may wish to consider when evaluating discovery services, concentrating on products which are deployed in the UK; there are others which are worthy of consideration such as Bibliocommons<sup>1</sup>, eXtensible Catalog<sup>2</sup>, and Visualizer<sup>3</sup> from VTLS.

As electronic resources began to grow as a proportion of collections, traditional OPACs became more limited as a means of directing users to the resources they required. Federated search, searching across a number of discrete databases from a single interface, became popular around the millennium but was only partially successful. Cross-searching databases with differently structured metadata produced patchy and incomplete results. Federated search may have been most useful in highlighting to the user which databases to search.

Around 5 years ago, as libraries were finding that traditional OPACs were not able to guide users to non-print materials and articles, and that federated search was not a satisfactory solution, 'web-scale' discovery services began to emerge, built around unified indexes of licensed scholarly publications combined with metadata for the local catalogue and other local content, with a Google-like simple search box, and a consolidated, single result list. In the intervening years they have become increasingly popular, particularly with Higher Education libraries, where it is common to deploy a discovery service separate from the OPAC, and often from a different supplier to the library's LMS. First to market was OCLC's WorldCat Local in 2007, followed by Serials Solutions' Summon in 2009, and in 2010 EBSCO Discovery Service (EDS), Innovative Interfaces' Encore Synergy and Ex Libris' Primo Central. All are considered below.

Library system vendors have updated OPACs in recent years to have a fresh look and feel and functionality similar to the discovery services, with faceted search limiting and web 2.0 features, and in some cases the ability to ingest other local data (such as digital repository data) alongside catalogue metadata. For example,

<sup>1</sup> <http://www.bibliocommons.com/>

<sup>2</sup> <http://www.extensiblecatalog.org/>

<sup>3</sup> <http://www.vtls.com/products/visualizer>

see Capita's Prism<sup>4</sup>, or SirsiDynix's Enterprise<sup>5</sup>. However these products remain hampered by their inability to offer access to licensed 3<sup>rd</sup> party or article level data. The 'web-scale' discovery services promote discovery across the whole library collection, not just catalogued material.

A question that should be asked of discovery layers is whether they are making use of RDA (Resource Description and Access) to enhance displays. RDA is organised based on FRBR (Functional Requirements for Bibliographic Records), "Identifying 'user tasks' which a library catalogue should make possible and a hierarchy of relationships in bibliographic data"<sup>6</sup> RDA supports clustering of bibliographic records ('FRBR-isation') to show relationships between works and creators, and make users aware of different editions, translations or physical formats.

There are broadly, 2 categories of 'discovery layer' distinct from the OPAC

1) powerful search interfaces on top of locally built indexes, (potentially with data from different sources), created and maintained by the library.

2) vendor-hosted indexes, including items the library has no local metadata for, but to which the library can add its own local data (the 'web-scale' discovery services).

The library has more control over category 1, and over how local metadata is indexed and searched, but category 2 is generally required to provide access to resources at article level. The leading products in both categories are considered below.

## Discovery interfaces on local databases

Blacklight - <http://projectblacklight.org/>

Blacklight is an open source product, not affiliated or coupled with any LMS<sup>7</sup>. It can provide a discovery interface for any index built using Apache Solr. The Solr index can include data from several sources, e.g. catalogue, digital repository or other collections.

Blacklight has a clean, modern interface which is customisable. It offers faceted browsing, relevance based searching (with the ability to locally control the relevancy algorithms, as opposed to the proprietary relevance

<sup>4</sup> <http://www.capita-softwareandmanagementservices.co.uk/software/pages/libraries-prism.aspx>

<sup>5</sup> <http://www.sirsidynix.com/enterprise>

<sup>6</sup> Oliver, Chris (2010). *Introducing RDA: a guide to the basics*. ALA Editions. pp. 128. ISBN 978-0-8389-3594-1.

<sup>7</sup> <http://projectblacklight.org/>

algorithms used by the web-scale discovery services considered below), bookmarkable items, permanent URLs for every item, and user tagging of items. There is a variety of filter options and sort options – language, format, location, topic etc. Availability of items can be looked up in real time from the LMS. Like the ‘web-scale’ discovery services considered below, Blacklight can present the user with formatted citations.

Blacklight is used by a large number of libraries in the USA, including University of Virginia, Stanford University, Johns Hopkins University, and WGBH (all of which have contributed to the product’s code).

In the UK, University of Hull uses Blacklight for their institutional repository - <http://hydra.hull.ac.uk> and also offer a Blacklight version of the library catalogue, <http://blacklight.hull.ac.uk/> alongside Hull’s Millennium OPAC. However as Blacklight cannot provide access to databases, full text articles etc, University of Hull also offer Summon.

[VuFind - http://vufind.org/](http://vufind.org/)

VuFind was developed by Villanova University and was released in July 2010. Like Blacklight, it is open source software and not affiliated or coupled with any LMS. VuFind is however marketed and supported in the UK by PTFS Europe, primarily as the discovery interface for the Koha LMS.

VuFind also uses a Solr index and can index locally sourced metadata from disparate sources, including but not limited to the library catalogue. It presents a simple, Google-like interface with keyword searching. It can also be deployed as an interface layer via APIs to search Summon, EDS or Primo.

VuFind is modular so it is up to the user how many features and functions to implement. Its open source nature allows technically minded users to modify modules or add new functionality.

Over 100 institutions are using or evaluating VuFind<sup>8</sup>. Institutions running live instances include the London School of Economics, the National Library of Ireland, Yale University, Birkbeck College and University of Kent.

VuFind offers limiting by facets including format, call number, language, author, and genre. Other features include suggested resources and searches, browsing, creating lists, texting, e-mailing, tagging, and commenting features, persistent URLs, and multi-language capability.

However, as with Blacklight, VuFind cannot index 3<sup>rd</sup> party data or provide access to databases, full-text articles etc.

<sup>8</sup> [http://vufind.org/wiki/installation\\_status](http://vufind.org/wiki/installation_status)

[Aquabrowser - http://www.serialssolutions.com/en/services/aquabrowser/](http://www.serialssolutions.com/en/services/aquabrowser/)

Another discovery layer which can harvest data from the library catalogue and other sources but does not offer an index of 3<sup>rd</sup> party database content is Aquabrowser from Serials Solutions.

Aquabrowser is a well-established product used by more than 700 libraries worldwide. It is offered as a hosted service.

Aquabrowser offers library-configurable relevancy ranking and faceted browsing, with unlimited facet categories. Item availability is retrieved from the LMS in real time. Aquabrowser can integrate with most LMSs, offering " the most extensive ILS support of any discovery layer"<sup>9</sup> and it also offers web 2.0 social features.

The Aquabrowser interface presents a tag cloud to the left of results, and facets to the right. The interface was revolutionary when first deployed nearly a decade ago, but is now somewhat dated. Few libraries are now deploying Aquabrowser for the first time as the 'web-scale' discovery services or open source rivals offer superior functionality.

### **Web-scale discovery services**

Vaughan<sup>10</sup> summarises what 'web-scale' discovery services offer as content, discovery, delivery, and flexibility (i.e. LMS and other system agnostic). Using any of these services the user can search beyond what is owned/licensed by the library - which presents both pros and cons.

All of the 'web-scale' services offer integrated chat widgets, icons to represent content types, full-text searching, relevancy ranking, limiting and sorting and permalinks. All allow search boxes to be embedded in other websites.

In all cases, cataloguing takes place in the LMS – metadata is not created directly in the discovery service. Hence, all the 'web-scale' discovery services treat metadata from different sources differently, which can cause issues with search and retrieval. Han<sup>11</sup> points out that to maximise functionality of discovery services, libraries should review current cataloguing practices to better serve users' needs.

<sup>9</sup> <http://www.serialssolutions.com/en/services/aquabrowser/features>

<sup>10</sup> Vaughan J. *Investigations into library web-scale discovery services. Information Technology and Libraries. 2012;31:32*

<sup>11</sup> Myung-Ja Han. *New Discovery Services and Library Bibliographic Control. Library Trends. 2012;61:162-172*

Summon - <http://www.serialssolutions.com/en/services/summon>

Serials Solutions' Summon is a hosted, cloud-based, multi-tenant system. Summon is used by over 500 libraries worldwide<sup>12</sup> and 16 in the UK<sup>13</sup>. It is not coupled with any LMS at present and can be used in tandem with many different systems. Summon does require regular catalogue data updates (as do the other discovery services, if deployed separately from an LMS), but can look up availability of holdings from the LMS in real-time. As with any of the discovery services if deployed separately from the LMS, the lack of integration with the LMS means users are not presented with a direct link to reserve/renew items, or check their account. However Serials Solutions is developing a new "Library Services Platform" or next-generation LMS, Intota, which is expected to be fully integrated with Summon.

The Summon service is built around a central index which includes, alongside local catalogue records and data harvested from other local sources (e.g. digital repositories), content from 3<sup>rd</sup> party suppliers of databases and e-journals. Serials Solutions highlights as unique to Summon the large amount of Abstract and Indexing database content, available to libraries subscribing to these databases.

The user is presented with a simple search interface and results are returned quickly, with facets to the left of the screen for narrowing search results.

If the library wants to offer a different interface to the default, Summon can integrate with Serials Solutions' own Aquabrowser, and also offers a suite of open APIs for use by third party vendors. A library can build its own interface using Summon's API or use an existing interface such as VuFind or Blacklight to search Summon.

Summon default interface is easy to use and intuitive. However, there are resources which are not indexed, and the service fails to link to some resources, particularly those provided by EBSCO. (Conversely, EBSCO's EDS is poor in providing access to resources provided by ProQuest, the parent company of Serials Solutions).

Summon works well on mobile devices. The refinement options from the left of the screen move to the top of the mobile screen quite neatly. However, third-party content linked to from Summon often does not display as well on a mobile screen. This is not a problem with Summon itself, and indeed this is also the case when using any of the other 'web-scale' discovery services on a mobile.

<sup>12</sup> [http://www.serialssolutions.com/en/press\\_room/detail/serials-solutions-delivers-new-contextual-research-assistance-in-the-summon](http://www.serialssolutions.com/en/press_room/detail/serials-solutions-delivers-new-contextual-research-assistance-in-the-summon)

<sup>13</sup> [http://community.summon.serialssolutions.com/index.php/Summon\\_Customer\\_List](http://community.summon.serialssolutions.com/index.php/Summon_Customer_List)

There is a new release of Summon approximately every 3 weeks. Unlike Primo, World Cat Local and EDS, it was built from scratch, not based on an existing product.

Summon is primarily marketed to academic libraries and a subscription is based on Full-Time Equivalent student count. Implementation typically takes 2-3 months, though can be done in as little as 6 weeks.

Unique to Summon is the 'database recommender'. Using this feature the library can promote databases whose full text is not available via Summon using keywords which when searched on by users can link to those databases.

Part of the Summon service is a statistics package including reports on number of searches performed, and commonly used search terms.

Summon's effectiveness depends on what type of resource the user is looking for, and whether they need 'just enough' information, or everything available on a topic. The needs of undergraduates can be contrasted with those of researchers; the latter will still need to use databases outside Summon.

[Primo Central - http://www.exlibrisgroup.com/category/PrimoOverview](http://www.exlibrisgroup.com/category/PrimoOverview)

Ex Libris' Primo Central, like Summon, is built around a unified index of aggregated data. From a single search box users can explore local collections, and 3<sup>rd</sup> party resources, presented in a single, relevancy-ranked list. Not all subjects are covered as well as others in the Primo Central index. Primo however can also perform federated searching in addition to searching the central index.

Primo can be used with any LMS but is integrated with the Ex Libris LMSs Aleph and Alma. Users in the EU are unable to use any other discovery layer with Alma. Primo Central harvests bibliographic data from Alma daily, and looks up holdings and availability in real time. Primo has over 1200 users worldwide<sup>14</sup>.

Primo Central is sold as a hosted, cloud-based service, though it can also be locally deployed. It is built on the Primo system which was launched in 2005. Primo is relatively expensive to licence, compared with competitor services. Pricing is based on Full-Time Equivalent students and number of local records indexed.

The Primo interface can be extensively customised; APIs allow libraries to more or less build their own interface.

Unique to Primo is the bX recommender service for scholarly articles, directing users to content other users have viewed based on what the user is looking at, built around usage logs.

<sup>14</sup> <http://www.exlibrisgroup.com/category/PrimoOverview>

Primo is easy to use, like Summon presenting a relevancy-ranked list of results that can be narrowed by selecting facets from the left-hand pane.

Primo also looks good and works well on mobile, but the mobile site does not offer the full range of functionality of the service.

Usage statistics are provided as part of the service. A new version of Primo is released quarterly.

[EBSCO Discovery Service - http://www.ebscohost.com/discovery](http://www.ebscohost.com/discovery)

EDS is not coupled with any LMS, and EBSCO has not announced plans to develop a library services platform. (However EBSCO is partnering with SirsiDynix, Capita and Innovative Interfaces to offer its centralised index to users of the LMSs provided by those vendors).

EBSCO's stated aim is to give as good a user experience and functionality as searching academic research databases<sup>15</sup>.

EDS is another hosted, cloud-based service. Subscription pricing is based on the institution's Full-Time Equivalent student count and level of service (number and types of local resources indexed). EDS offers content from nearly 20,000 providers.

The library can customise the interface to a degree, including the institution's own corporate colours and logos.

EDS offers a single search box to search the library catalogue and databases. The interface is based on the long-established EBSCOHost platform (launched 1994). Content includes Netlibrary eBooks. Link resolution is not required for EBSCO content, though it is for non-EBSCO content.

Results are ranked by relevancy by default and result sets can be limited using facets in the left pane. Unlike the other 'web-scale' services, EBSCO explains the relevancy algorithm used on its support website. Since 2011, EDS has offered a level of FRBR-isation in search results.

Additional resources to search can be added in the right pane (federated search). Other features include RSS feeds and email alerts, spell checking and citation formats. Results are returned quickly.

In my testing I found EDS was not optimised for mobile screens as well as some of the rival services.

Perhaps because it aims to meet the needs of postgraduate and academic researchers as well as the simpler searching needs of undergraduates, EDS is not so intuitive and easy to use as Summon or Primo, and as noted above, access to ProQuest databases is poor.

<sup>15</sup> <http://www.EBSCOhost.com/discovery/about>



EBSCO claims to offer superior metadata to the rival 'web-scale' services, and also to offer more searchable full-text. EBSCO claims "the overwhelming majority of the most respected indexes do not provide their metadata to any discovery service"<sup>16</sup>

Statistical reports are provided with EDS. For EBSCOhost databases, these include number of sessions and searches, number of abstract views and number of full-text article requests.

An EDS implementation typically takes 8-10 weeks. EDS may be a good choice for libraries which are already familiar with, and subscribing to content on the EBSCOHost platform, with which EDS is integrated. As with the rival services, there are gaps in coverage in some subject areas, e.g. law, market research, but the coverage is improving (also as with the rival services).

[WorldCat Local - http://oclc.org/en-UK/worldcat-local.html](http://oclc.org/en-UK/worldcat-local.html)

OCLC's WorldCat Local is built around the WorldCat database (as of 18 Jan 2013 containing bibliographic data for 775m articles and 233m books). WorldCat Local also provides access to OCLC licensed full-text databases via WorldCat, with a federated search component for non-OCLC licensed databases.

WorldCat Local interoperates with local LMSs for retrieval of real-time item availability, and is fully integrated with OCLC's WorldShare LMS. OCLC's link resolver is built in to WorldCat Local.

Like the other 'web-scale' services, WorldCat Local offers relevancy ranking of results, facets for limiting result sets, enriched content including book jacket covers and descriptions/contents pages, web 2.0 social features and links to Amazon and Google Book Search.

WorldCat Local does FRBR-ise results, but reaction to this is mixed. Editions and formats are grouped together but the most popular rather than the latest edition is shown in the summary of results.

WorldCat Local searches all of WorldCat (unless scoped to only search locally) and sends local search results to the top of the list. Some resources are promoted better than others in WorldCat Local search results. The centrally indexed WorldCat record is displayed and not the library's local bibliographic record. Indeed WorldCat Local works best if a library uses OCLC for cataloguing; otherwise frequent batch loads of catalogue records are required to keep the service synchronised with the local catalogue.

WorldCat Local is priced lower than the rival 'web-scale' services (pricing based on library user population). It is used by 399 libraries worldwide but presently by fewer than 10 libraries in the UK. Unlike the rival services, WorldCat Local is marketed at all types of libraries rather than focussing on the academic sector. A

<sup>16</sup> <http://www.ebscohost.com/discovery/technology/platform-blending>

typical WorldCat Local implementation takes 2-3 months. A suite of statistical reports is provided, and the mobile version of the service works well.

There are quarterly WorldCat Local releases. The WorldCat Local product is currently undergoing a re-architectchure and the new version will be released in early 2014.

[Encore Synergy - http://encoreforlibraries.com/tag/encore-synergy/](http://encoreforlibraries.com/tag/encore-synergy/)

Encore Synergy from Innovative Interfaces is integrated with the Millennium and Sierra LMSs, but can be used with other systems.

Encore Synergy offers faceted search, and relevancy ranking, together with social features.

Facets are presented to the left of the screen and optionally a tag cloud and more services to right of results (this can make for a cluttered screen). In my testing Encore Synergy did not work as well on a mobile screen as some of the other discovery services, simply displaying the standard interface on the smaller mobile screen.

Unlike Summon, Primo Central, EDS and WorldCat Local, Encore Synergy does not offer an integrated catalogue and article search – they need to be searched separately, though in the same interface.

Encore Synergy has no centrally aggregated index but instead accesses publishers and aggregators in real-time through web services, a form of federated searching.

## **Conclusions and looking to the future**

According to Luther and Kelly<sup>17</sup> discovery services can be evaluated on content (scope and depth, richness of metadata, frequency of updates), search (simplicity of interface, quality of results etc), fit (ease of implementation, compatibility with existing systems, responsiveness etc of vendor), and cost.

Interoperability with the LMS/OPAC is an important consideration in selecting a discovery service. However the OPAC as a separate entity may be dying; neither Alma nor Intota offer an OPAC distinct from the discovery service. As Breeding puts it, traditional OPACs are usually still made available alongside the discovery service because "I continue to hear voices of dissent from not just librarians, but also from groups of library users with more complex requirements"<sup>18</sup>. However these traditional OPACs are getting older all the time and not being actively developed. Lorcan Dempsey said in 2012 "There is a renaissance of interest in the catalog and catalog data. Yet it comes at a time when the catalog itself is being reconfigured in ways

<sup>17</sup> Luther, J., & Kelly, M. C. (2011). *The next generation of discovery: The stage is set for a simpler search for users, but choosing a product is much more complex.* (LJ SERIES: THE BIG TOOLS). *Library Journal*, 136(5), 66.

<sup>18</sup> Breeding, M. *Looking Forward to the Next Generation of Discovery Services.* *Computers in Libraries [serial online]*. 2012;32:28.

which may result in its disappearance as an individually identifiable component of library service"<sup>19</sup>. Dempsey cites facets and FRBR-isation as ways to "make the data work harder".

Technical support for the discovery service is another consideration. Open source tools may need more management/configuration/technical expertise (though there are companies which offer support at a cost). The proprietary services depend on the service provided by the vendor, which may be as relevant a consideration as functionality. Libraries may also be looking for an active user community - all of the services considered above have an active email list where users will support each other.

The differences in content between the services should be considered when selecting a service, but they are becoming less pronounced as the vendors strike more deals with content providers. It should be borne in mind that there will always be some content not indexed by any of the discovery services, and for that reason, some libraries may still see a role for some degree of federated search.

The library should always ask, "how does a discovery service enable your users to find the resources they need?" Focus on the user experience is critical - even if it breaks professional standards.

In conclusion, Paul Stainthorp is correct in his assessment of the 'web-scale' services that "the differences between [them] are *not that significant*...thinking that ...there are some 'good' and some 'bad'...is probably wrong. It's not really about the product, it's about the willingness of the vendor to overcome problems, and about their attitude to their customers"<sup>20</sup>. Stainthorp also highlights de-duplicating via FRBR and known-item searching as weaknesses common to all services.

A difficult task but one that should be attempted is to make sure a solution is future-proof. Libraries should monitor vendors' development plans for discovery services. Ken Chad<sup>21</sup> advises reference to Gartner's technology trends for 2013<sup>22</sup> which include moving to mobile, moving to the cloud, actionable analytics, and strategic Big data (aggregation of data from discrete sources).

Looking to the future, particularly regarding metadata, it will be interesting to see the results from NISO's Open Discovery Initiative which "aims at defining standards and/or best practices for the new generation of library

<sup>19</sup> <http://www.educause.edu/ero/article/thirteen-ways-looking-libraries-discovery-and-catalog-scale-workflow-attention>

<sup>20</sup> <http://paulstainthorp.com/2011/05/17/how-commercial-next-generation-library-discovery-tools-have-nearly-got-it-right/>

<sup>21</sup> <http://www.slideshare.net/kenchad/jibs-discovery-servicesfeb2013kenchad>

<sup>22</sup> <http://www.gartner.com/newsroom/id/2209615>

discovery services that are based on indexed search.”<sup>23</sup>

There are signs in the market of a move to integrated systems, bringing the LMS and discovery services together (e.g. Alma and Primo, Intota and Summon, WorldShare and WorldCat Local; none of these LMSs will work with a discovery service from another provider). From the vendor’s point of view this offers more control of the solution space and systems that are easier to support. From the library’s point of view an integrated solution could offer cost savings and simplified configuration and workflows. This is a developing situation and it is difficult and may be dangerous to predict medium or long-term trends.

<sup>23</sup> <http://www.niso.org/workrooms/odi/>