

# Web Scale Discovery Service and Tools with Special Reference to Central Library, IIT Kharagpur

Parnab Chatterjee<sup>1</sup>, Debasis Das<sup>2</sup>

<sup>1</sup> Librarian, Janardandih High School, Purulia- 723144, India, Email:parnabchatterjee@gmail.com

<sup>2</sup> Library Professional, Central Library, IIT Kharagpur- 721302, India, E mail:debdas12312@gmail.com

## Abstract:

The purpose of this paper is to examine and compare discovery services and tools. The paper summarizes the characteristics discovery services and tools into a check-list of 12 features. This paper also checked the major comparison between Proprietary Discover tools. The paper also discusses the service provided by Central Library of IIT Kharagpur i.e. “eSearch Software”. Discovery tools have many next generation catalog features, but only a few can be called real next generation catalogs. Federated searching and relevancy based on circulation statistics are the two areas that both open source and proprietary discovery tools are missing. This paper gives information about discovery service provide by Central Library of IIT Kharagpur. It will provide information as to exactly where discovery tools stand in light of the much desired next generation catalog.

**Keywords:** Web Scale Discovery Services, Discovery Tools, Federated search, Next-Generation Cataloguing, eSearch Software.

## 1. Introduction:

‘Discovery services’ is a new name for the traditional information retrieval system. A bit of market confusion is created by the new discovery product, vendors, projecting their products as a potential substitute for many search services that the libraries have been using for decades like card catalogue, OPAC, Web Opac, Bibliographic database search etc. Two things are projected as key benefits of these new age discovery services. First they act as a single point search interface for all the contents (subscribed, in-house and open access) that the library wants to present to its user community for searching. It is mimicking Google, and second they operate on web scale. Most of libraries have different types of electronic resources and each e-resources database has their own search engine therefore users need to visit different search engines to get their required information for which it is required services a single point search interface for all the content has emerged, which is more effective and useful for the users.

Web scale discovery services are those integrated web based services with major potential to transform the nature of library systems. These services are offered as cloud computing model and have the capacity to more easily connect researchers with the library's vast information repository including remotely hosted resources and local content. It

provides a unified platform for library users to access and search from all the library resources to get single set of results by providing a Google like environment with the following basic features.

- Unified platform to search all the resources including licensed, open and local collections
- Pre-harvested central index of metadata
- Google like single search box
- Single results list for all collections
- Relevancy ranking across entire results
- Full featured user interface
- Facets and tools for narrowing results
- Holdings and status information for library catalogue items
- Connections to full text
- Infrastructure, processing and indexing provided and maintained remotely by the vendor.

A discovery tool is often referred to as a stand-alone OPAC, a discovery layer, a discovery layer interface, an OPAC replacement, or the next generation catalog (NGC).

Unlike the front end of an integrated library system or ILS OPAC, a discovery tool is defined as a third party component whose purpose is to “provide search and discovery functionality and may include features such as relevance ranking, spell checking, tagging, enhanced content, search facets” (OLE Project, 2009). Discovery tools should not be confused with federated search products. The former “promise to provide a single interface to multiple resources based on using a centralized consolidated index to provide faster and better search results”, while the latter search remotely, rely on connectors, and provide “only partial and limited solutions” (Hane, 2009). In addition, a federated search tool usually requires user logon and works in a protected environment, while a discovery layer is open to the public. A federated search tool is dedicated to finding articles across a number of subscribed databases and as such is not within the scope of this paper. Libraries are disappointed with commercial ILS OPACs. Developed as a part of an integrated library system, they have remained relatively static over the years and have not evolved in pace with the discovery and search tools now commonplace at commercial sites such as Amazon. Most of them cannot and will never be able to provide advanced functionalities in order to meet current expectations. It is more practical for vendors and developers to field new OPAC systems that run alongside the older ones than to attempt to alter the proprietary code of ILS OPACs. Most current ILS OPACs do not offer the features of these standalone, next generation catalogs.

Until recently, libraries could do nothing about their outdated OPAC. Proprietary ILS OPACs offered only limited customization. Today, libraries using some of the ILS OPACs can add patches and a limited number of functional improvements by acquiring both free and commercially available plug-ins and add-on modules, but this solution will not completely transform an old OPAC into a next generation catalog.

Additionally, libraries may adopt a “Web OPAC wrapper” solution to embed their existing OPAC within another user interface layer (Murray, 2008). The current trend some libraries seem to favor is to simply abandon their current OPAC in favor of one of the new standalone, next-generation discovery tools.

#### Discovery tool evaluation check-list:

Yang, S.Q and Wagner, K in their article “Evaluating and comparing discovery tools: how close are we towards next generation catalog?” summarizes characteristics of the next generation catalog into a check-list of 12 features. The features are given below:

1. Single point of entry for all library information. The library catalog should be a single search or federated search for all library materials, including pointers to the articles in electronic databases as well as records of books and digital collections. One search should retrieve all relevant materials. Presently, patrons have to search the catalog for books and videos, databases for journal articles, and digital collections and archives for local images and materials.
2. State-of-the-art web interface. Library catalogs should have a modern design similar to commercial, e-business sites. This criterion is highly subjective and as such is difficult to quantify. A next-generation catalog should look and feel like popular sites such as Google, Netflix and Amazon.
3. Enriched content. Library catalogs should include book cover images, user driven input such as comments, descriptions, ratings, and tag clouds. Traditionally, only professionally trained cataloging librarians have the ability to create or add content to bibliographical records.
4. Faceted navigation. Library catalogs should be able to display the search results as sets of categories based on some criterion such as dates, languages, availability, formats, locations, etc. Users can conduct a very simple, initial search by their preferred keyword method and then refine their results by clicking on the various results facets.
5. Simple keyword search box on every page. The next generation catalog starts with a simple keyword search box that looks like that of Google or Amazon. A link to advanced search should be provided. The simple search box should appear on every page of the interface as users navigate and conduct searches. Librarians prefer an advanced search and feel that the quick search is more likely to produce results with less precision.
6. Relevancy. Librarians complain that OPAC relevancy results are problematic or that they do not understand how relevance is determined. The next-generation catalog does better in relevancy ranking with increased precision. In addition circulation statistics should influence the relevancy results. More frequently circulated books indicate popularity and usefulness. They should be ranked higher in the display. Items deemed important enough to have multiple copies should also receive higher relevancy ranking.
7. Did you mean . . .? A spell-checking mechanism should be present in a next-generation catalog. When an error appears in the search, there should be a pop-up with the correct spelling or suggestions from a dictionary. Clicking on any of these runs a search.
8. Recommendations/related materials. Commonplace in e-commerce sites, the customer is shown additional items with a suggestion like “Customers who bought this item also bought . . .” Likewise; a next-generation catalog should recommend books for readers on transaction logs. This should take the form of “Readers who borrowed this book also borrowed the following . . .” or a link to “Recommended Readings”.

9. User contribution. The next-generation catalog allows users to add data to records. The user input includes descriptions, summaries, reviews, criticism, comments, rating and ranking, and tagging or folksonomies. Today's users increasingly look for what other users have to say about items found online, and value what they feel to be their peers' review of items. Tagging clouds can serve as access points and descriptive keywords leading to frequently used items.
10. RSS feeds. Really Simple Syndication allows users to connect themselves to content that is often updated. Next-generation interfaces include RSS feeds so that users can have new book lists, top-circulating book lists, canned searches, and "watch this topic" connections to the catalog on their own blog or feed reader page.
11. Integration with social network sites. When a library's catalog is integrated with social network sites, patrons can share links to library items with their friends on social networks like Twitter, Facebook and Delicious.
12. Persistent links. Next-generation catalog records contain a stable URL capable of being copied and pasted and serving as a permanent link to that record.

## **2. Major discovery products providers: Open source and proprietary**

Sharon Yang and Kurt Wagner's presentation on open source discovery tools at the Virtual Academic Library Environment (VALE) 2010 Annual Conference was used to identify these products (Yang and Wagner, 2010).

Open source discovery tools are considered separately from commercial, proprietary products for the simple reason that the former can be freely implemented, customized and used. They require some local programming and configuration to enable them to search and display data from a traditional ILS. These open source products do not require any sort of contract, or support, as is the case with proprietary systems. The second list is for evolved, next-generation interfaces offered by commercial ILS or interface vendors. The following are two alphabetical lists of sites, one for open source and one for proprietary discovery tools reviewed in this study:

### **2.1 Open Source Discovery Tools**

1. Blacklight
2. Fac-Back-OPAC (Kochief)
3. LibraryFind
4. Rapi
5. Scriblio (WPopac)
6. SOPAC (Social Opac)
7. VuFind

### **2.2 Proprietary Discovery Tools**

1. Aquabrowser by Serials Solutions
2. BiblioCommons

3. EBESCO Discovery Service
4. Encore-Innovative Interfaces Inc.
5. Endeca-EndecaOne Search: Follett
6. Primo-Ex Libris
7. SirsiDynix Enterprise-SirsiDynix
8. Summon by Serials Solutions (now Proquest)
9. Visualizer-VTLS
10. WorldCat Local-OCLC

### 2.3 Comparison of Discovery services

The effectiveness and efficiency of discovery services are based mainly on two factors. One is content coverage and the other one is technology aspects utilized in various sub systems like harvesting, searching, relevancy ranking, interface features etc. Discovery solutions provided by various service providers have varying degree of differentiations in these features. A comparison of four 3 major commercial cloud based discovery services is made based on some important parameters which are decisive in the choice for the customers.

**Table 1: Comparative analysis of major service providers -**

#	Summon	Primo Central	WorldCat Local
<b>Vendor</b>	Serial Solution	Exlibris	OCLC
<b>License</b>	Proprietary	Proprietary	Proprietary
<b>Hosting/Installation</b>	hosted	hosted/UI may local	Hosted
<b>Support</b>	From Vendor	From Vendor	From Vendor
<b>Central index</b>	Hosted	Hosted	Hosted
<b>Harvesting</b>	From open source & commercial)	From open source & commercial	From open source & commercial
<b>Relevance ranking</b>	based on proprietary algorithm	based on proprietary algorithm	based on proprietary algorithm

<b>Save search items</b>	√	√	√
<b>Catalogue item availability indication</b>	√	√	√
<b>Refine result by categories</b>	√	√	√
<b>Faceted Display of result</b>	√	√	√
<b>Support mobile devices</b>	√	√	√
<b>Did you mean suggestions</b>	√	√	√
<b>RSS feed</b>	√	√	×
<b>Multiple language interface</b>	√	√	√
<b>Price</b>	FTE and local collections	FTE and local collections	FTE and local collections
<b>Customization (branding, colours)</b>	customizable	customizable	customizable
<b>Providing custom links (eg:- library site)</b>	customizable	customizable	customizable
<b>Custom URL for WSD</b>	×	×	√
<b>Search box can be in external sites</b>	√	√	√
<b>Customer can supply CSS</b>	×	√	×
<b>RSS</b>	√	√	√
<b>Export to reference tools</b>	√	√	√

<b>User ratings, user reviews, user tags</b>	x	√	√
<b>Tag clouds</b>	x	√	x
<b>Platform Blending</b>	x	x	x
<b>Single point of entry</b>	x	√	x
<b>State-of-the-art web interface</b>	√	√	√
<b>Enriched content</b>	√	√	√
<b>A simple box of keyword search with a link to advanced search</b>	x	√	√
<b>Recommendations</b>	√	x	√
<b>Integration with social networking sites</b>	√	x	x
<b>Persistent link</b>	x	√	x

This comparative analysis shows that all the major service providers are extending competitive features and services, but varies in some features and the choice is depends on the concerned library’s preferences and the cost involved.

### 3. Implementation of eSearch Software for providing Discovery Services at IIT Kharagpur

The Central Library of IIT Kharagpur started in the year 1951, and moved to its present premises in 1956. For the last six decades, the library has been the lifeline of the academic activities of the Institute. It is no exaggeration to state that the Central Library of IIT Kharagpur is one of the largest and finest technical libraries in Asia. The Library is well equipped with modem facilities and resources (print and electronic) in the forms of CDROMs, DVD-ROM, online databases, micro-documents, video cassettes, books, journals, patents, standards, theses, reports, etc. It has also developed a full-fledged Digital Library equipped with necessary modem equipment in order to provide various digital mode Library services. Digital Library section of the Central Library has developed an institutional digital repository using open source software DSpace, which preserve the institute's intellectual output for campus wide access and digital preservation for the posterity.

The central library started to design and develop e-Search Software for providing Discovery Services in the year 2010. e-Search Software is a single window search box to access e-resources. Library made an evaluative study of various user interface of e-resource search software mainly MIT “VERA”, Stanford University “e-Journal Search” and Stanford University “Search Works”.

After evaluative study of e-Search software Central Library decided to develop in-house software similar to VERA using open source software. The next version of this software will be developed with a feature to search full text.

### 3.1 Feature of e-Search Software

1. Platform Independent: Server Search Engine may be installed in any type of computer.
2. Develop Using FOSS
3. eJournal Title Search
4. eBook Title Search
5. Boolean Title Search: Present in Advanced Search mode
6. Full Text Search: Will be available in the next version
7. Open Access Journals are Included
8. Use of Wild Cards: % is used to match any number of characters:
9. Integration of other Social Tool or Databases: SCOPUS, WOS, Library Web Opac, Standards etc. are included and gets opened in the same container application.
10. Feedback Mechanism

The eSearch software is web application software therefore any well configured computer will support to install eSearch software. Most of the Feedback received from the faculty and the students are mentioning that the eSearch is very handy and useful. The only shortfall of the software is that it does not perform any full text search which is needed by today’s information seekers. The forth coming version of this software will have a full text search facility. Those library users who are searching books on title name this tool is very helpful for them. The search engine performs a search from the local database; therefore, the search result is very quick and precise. Presently, The Central Library, IIT Kharagpur provides Discovery services, but this is a trial access. Central Library of IIT Kharagpur provides Informatics-Federated Discovery Service, Ebsco- Discovery Service and Summon- Discovery Service.

### 4. Conclusion

Web Scale services are still in its initial stages of development and lots of developments in the features, functionality, level of integration with other systems, scope of content, and soundness of metadata, flexibility of the interface are all evolving and it is expected, will continue to evolve in meeting the needs and expectations today’s next generation users. The comparative analysis shows that all the major service providers are extending competitive features and services, but varies in some features and the choice is depends on the concerned library’s preference

and the cost involved. A discovery tool is not complete without this federated search capability. True federated searching is, and will always be, the promised land of next-generation catalog and discovery tools. Libraries, vendors and the open source community must continue to cooperate and work in a spirit of optimism and collegiality to make the true next-generation catalog a reality.

Selecting, evaluating and implementing a web-scale discovery products has taught us much about project effectiveness, communication strategies, implementation processes and ongoing challenges. Such lessons will stand us in good stead not only in this instance but also in the future as we continue to grapple with the ever-increasing rate of technological change and innovation.

## References:

1. Anita K, F., & Popp, M. P. (Eds.). (2012). "Early Adoption: EBSCO Discovery Service at Illinois State University: Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/early-adoption-ebSCO-discovery-service/67838>
2. Asher, A. D., Duke, L. M., & Wilson, S. (2012). "Paths of Discovery: Comparing the Search effectiveness of EBSCO Discovery Service, Summon, Google Scholar, and Conventional Library Resources". College & Research Libraries. Retrieved from <http://crl.acrl.org/content/early/2012/05/07/crl-374.short>
3. Breeding, M. (2010). "The State of the Art in Library Discovery." *Computers in Libraries*, 30(1), pp. 31–34.
4. Breeding, M. (2013). "Beyond the ILS: A New Generation of Library Science Platforms". IGI Global. Retrieved From [www.igi-global.com/chapter/beyond-ils-new-generation-library/76457](http://www.igi-global.com/chapter/beyond-ils-new-generation-library/76457)
5. Caplan, P. (2012). "On discovery tools, OPACs and the motion of library language". *Hi Tech*, 30(1), pp. 108–115
6. Comeaux, D. J. (2012). "Usability Testing of a Web-Scale Discovery System at an Academic Library". <http://dx.doi.org/10.1080/10691316.2012.695671>
7. Dallis, D. (Ed.), "Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/evaluating-selecting-library-web-scale/67814>
8. David, B., & Popp, M. P. (Eds.). (2012). "Designing an Evaluation Process for Resource Discovery Tools. Planning and Implementing Resource Discovery Tools in Academic Libraries". IGI Global. Retrieved from <http://www.igi-global.com/chapter/designing-evaluation-process-resource-discovery/67818>
9. Fagan, J. C., Mandernach, M., Nelson, C. S., Paulo, J. R., Saunders, G. (2012). "Usability Test Results for a Discovery Tool in an Academic Library". *Information Technology and Libraries*. Retrieved From [ejournals.bc.edu/ojs/index.php/ital/article/view/1855](http://ejournals.bc.edu/ojs/index.php/ital/article/view/1855)
10. Fagan, J.C. (2012). "Top 10 Discovery Tool Myths." *Journal of Web Librarianship*, 6(1), pp. 1–4.
11. Graves, T., & Dresselhaus, A. (2012). "One Academic Library: One Year of Web-scale Discovery:." *Serials Librarian (SERIALS LIBR)*.

12. Guthrie, A., and McCoy, R. (2012). "A Glimpse at Discovery Tools within the HBCU Library Landscape." *College & Undergraduate Libraries*, 19(2–4), pp. 297–311.
13. Hoepfner, A. (2012). "The Ins and Outs of Evaluating Web-Scale Discovery Services." *Computers in Libraries*, 32(3), pp. 6–10, 38–40.
14. Hoy, M. B. (2012.). "An Introduction to Web Scale Discovery Systems". *Medical Reference Services Quarterly (MED REF SERV Q)*. Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/02763869.2012.698186>
15. Kaufmann, K., Larsen, J., and DeSalvo, P. (2012). "Discovering the Discovery Tool: The Introduction and Impact on Research and Instruction at Seminole State College of Florida." *College & Undergraduate Libraries*, 19(2–4), pp. 278–296.
16. Leebaw, D., College, C., Conlan, B., College, S. O., Sinkler-Miller, C., College, C., Wilson, N., et al. (2012). "Improving Library Resource Discovery: Exploring the Possibilities of VuFind and Web Scale Discovery in a Consortial Environment". *Library Technology Conference*. Retrieved from [http://digitalcommons.maclester.edu/cgi/viewcontent.cgi?article=1263&context=libtech\\_conf](http://digitalcommons.maclester.edu/cgi/viewcontent.cgi?article=1263&context=libtech_conf)
17. *Library Technology Guides*. Retrieved from [librarytechnology.org](http://librarytechnology.org).
18. Narayanan, N., & Mukundan, R. (n.d.). *Cloud web scale discovery services landscape an overview*. Retrieved March 15, 2015, from <http://www.slideshare.net/.../cloud-web-scale-discovery-services-landscape-an->
19. Sutradhar, B. & Guha Roy, S. (2014). "Design and Development of e-Search software for providing Discovery services at Central Library, IIT Kharagpur"( Presented at the 9<sup>th</sup> Technology Book Fair & Workshops. Indian Institute of Technology, Kharagpur)
20. Vaughan, J. (2011). "Web Scale Discovery" *What Is Web Scale Discovery? Why Web Scale Discovery? Library Technology Reports: Expert Guides to Library Systems and Services*, 47(1), pp. 5–11.
21. Vaughan, J. (2011). *Web scale Discovery Services. Library Technology Reports*, 47(1), pp. 5–11.
22. Vaughan, J., & University of Nevada, L. V. (2011). "Investigations into library web scale discovery services". Retrieved from [http://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1043&context=lib\\_articles](http://digitalscholarship.unlv.edu/cgi/viewcontent.cgi?article=1043&context=lib_articles)
23. Way, D. (2010). "The Impact of Web-Scale Discovery on the Use of a Library Collection." *Serials Review*, 36(4), pp. 214–220.
24. Yang, S.Q., and Wagner, K. (2010). "Evaluating and Comparing Discovery Tools: How Close Are We Towards Next Generation Catalog?" *Library Hi Tech*, 28(4), pp. 690–709.