



Enhancing Access to Library Resources:

A Critical Discussion, Examination and Evaluation of Web-Scale
Discovery Services

Dr. Tariq Ashraf

University Librarian & Head

Dr. Zakir Husain Library

Jamia Millia Islamia (Central University), New Delhi

Email: tariq@south.du.ac.in

Agenda

- Why Discovery?
- Discovery Service....What?
- What it Does....?
- Going Ahead....Parameters

What is Discovery System?

- Libraries are essentially Resource Discovery Systems
- Libraries use different methods for resource discovery
- Library Discovery Systems have been changing from time to time
- Libraries have to provide more robust Discovery Services due to competition from online platforms.

Demand Declining for Traditional Library Services

- Defining Discovery Services
- Very few students and a decreasing number of faculty start research in the library building or via the library website, opting instead for search engines and discipline-specific research resources.
- Circulation and reference requests have been steadily declining for years, driving the library's traditional core (providing access to books and guiding patrons through research) to the periphery.

Viabile Alternatives to the Library Now Boast Fastest Growth and Easiest Access

- With the rise of companies like Google and Amazon, as well as non-profits like Wikipedia and HathiTrust, users now meet most of their information needs through sources outside of the library.
- Information available from multiple sources.

Free information on the web: Impact on the students

- Libraries no longer sole repositories of information.
- Web seen as easier, more accessible than library databases.
- Students have multiple options.

Resource Discovery Systems

- Library Catalogue
- OPAC
- WEB OPAC.
- Federated Search Engines
- MARC
- Web Scale Discovery

The 1st Discovery service (~1901)



Cataloging at LC, 1910/1920



MARC & OPAC (1980-1995)

- 1968 – 50,000 MARC records in existence
- 1982 – Card catalogs disappear
- OPACs appear in the library

Pros

- Users can search the entire physical collection quickly
- Tight ILS integration
- One place to search

Cons

- Users can only search catalog
- Metadata searching only

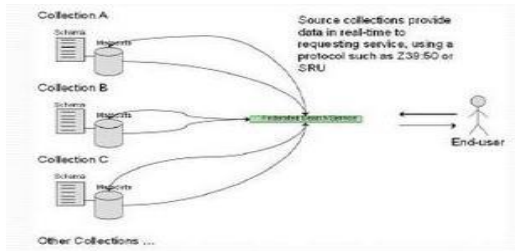
Federated Search

Pros:

- Single search box for all content
- Currency of content

Cons:

- Many indexes
- Multiple ranking algorithms
- Larger result sets incomplete
- Internet Traffic
- Content provider traffic



Catalog layers

Pros:

- Single index for all local content
- Speed
- Complete result sets

Cons:

- Usually combined with fed. search for e-content
- Different setup and software updates
- Not tightly integrated with ILS

Web-Scale Discovery

Pros:

- Search all content, one place
- Single search box
- Single index
- Single relevance ranking
- Speed
- Bandwidth
- Complete result sets
- No local hardware
- No local install

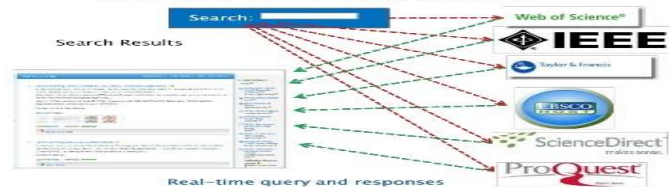
Cons:

- Not tightly integrated with ILS
- Easier setup (vendor maintains index)

Why Discovery Services

- Discovery services have the power to revolutionize library systems. With discovery services, you can find and retrieve information in ways never before possible.
- But what does this really mean for the average librarian? How do discovery services work? And finally, which service?
- What are the parameters for selection?

Federated Search Softwares



Discovery Services



Why Discovery Services? cont'd...

- As illustrated by research dating back primarily to the 1990s, library discovery systems within the networked online environment have evolved, yet continue to struggle to serve users.
- Library (or systems supported and maintained by the library) is often not the first stop for research – or worse, not a stop at all. Users accustomed to a quick, easy, “must have it now” environment have defected, and research continues to illustrate this fact.

Why Discovery Services? cont'd...

- Today, there are numerous alternative avenues for discovery, and libraries are challenged to determine what role they should appropriately play.
- Basic scholarly information use practices have shifted rapidly in recent years, and as a result the academic library is increasingly being dis-intermediated from the discovery process, risking irrelevance in one of its core functional areas.

Why Discovery Services? cont'd...

- It is our responsibility to assist our users in finding what they need without demanding that they acquire specialized knowledge or select among an array of “silo” systems whose distinctions seem arbitrary . . . the continuing proliferation of formats, tools, services, and technologies has upended how we arrange, retrieve, and present our holdings.
- Our users expect simplicity and immediate reward. Amazon, Google, and iTunes are the standards against which we are judged. Our current systems pale beside them.

Web Scale Discovery: What?

- Combining vast repositories of contents with accessible, intuitive interface, hold the potential to greatly facilitate the research process.
- Web-scale discovery services are able to index a variety of contents, whether hosted locally or remotely. Such contents can include library ILS records, digital collections, institutional repository contents, and contents from locally developed and hosted databases.
- Pre-index remotely hosted contents, whether purchased or licensed by the library. This latter set of contents – hundreds of millions of items – can include items such as e-books, publisher or aggregator contents for thousands of full-text journals, contents from abstracting and indexing databases, and materials housed in open-access repositories.
- Extremely rapid search and return of results ranked by relevancy, which can then be sorted in various ways according to the researcher's whim.

What Does Discovery Service Do?

- Creates a unified search interface for users pulling together information from the library catalog as well as other resources (e.g., journal articles, images, archival materials).
- Enhances discoverability of as broad a spectrum of library resources as possible.
- Intuitive: minimizes the skills, time, and efforts needed by our users to discover resources.
- Supports a high level of local customization (such as accommodation of branding and usability considerations).
- Supports a high level of interoperability (easily connecting and exchanging data with other systems that are part of our information infrastructure).
- Demonstrates commitment to sustainability and future enhancements.

Going Ahead: Some Parameters

- Identify vendor next generation discovery platforms, whether established and currently on the market, or those publicized and at an advanced stage of development, with an expectation of availability within a year's time. Identify & create a representative list of other academic libraries which have implemented or purchased currently available products.
- Create a checklist/ criteria of functional requirements/ desire for a next generation discovery platform.
- Create a list of questions to distribute to potential vendors and existing customers of next generation discovery platforms.
- Additional questions, such as pricing, maintenance, install base, etc.

Questions

- Seek to understand how contents hosted in our current online systems (Catalog, locally created databases, vendor databases, etc.) could/would (or not be able to) be incorporated or searchable within the discovery platform.
- How new information resources could be incorporated into the discovery platform.
- What types of existing records are discoverable within the vendor's next generation discovery platform, and seek an understanding of what basic metadata must exist for an item to be discoverable.

SEARCH

- The system has a “Google-like” simple search box
- The system has a “Google-like” simple search box, but also has an advanced search capability (user can refine the search to certain categories: author, journal, etc.)
- In case of zero results, the system should suggest something else or ask, “Did you mean?”

De-duplication of Similar Items

- The system automatically de-dupes records (the item only appears once in the returned list)

Web 2.0 Functionality

The following items are important for a discovery platform to have on results:

- A tag cloud
- Faceted searching
- Ability to add user-generated tags to materials ('folksonomies')
- Ability for users to write and post a review of an item

Enriched Record Information on Returned Results

- The following items are important to have in the discovery system:
 - Book covers for items held by the Libraries
 - A Google Books preview button for print items held by the Libraries
 - Displays item status information for print items held by the Libraries (example: available, checked out)

What the User Can do With the Results

- Retrieve the full-text of an item with only a single click on the item from the initial list of returned results
- Ability to add items to a cart for easy export (print, email, save, export to ref. works)
- Ability to place an Inter-Library Loan/ LINK+ Request for an item
- System has a login/ user account feature which can store user search information for later. In other words, a user could potentially login to retrieve saved searches, previously stored items, or create alerts when new materials become available.

Publisher/Aggregator Agreements: General

- With approximately how many publishers have you forged contents agreements with?
- Are these agreements indefinite or do they have expiration dates?
- Have you entered into any exclusive agreements with any publishers/ aggregators (i.e., the publisher/ aggregator is disallowed from forging agreements with competing discovery platform vendors, or disallowed from providing the same deep level of metadata/ full-text for indexing purposes).

Comments on Metadata Provided

- Could you please provide some general comments on the level of data provided to you, for indexing purposes, by the ‘majority’ of major publishers/aggregators with which you have forged agreements. Please describe to what degree the following elements play a role in your discovery service:
 - ‘Basic’ bibliographic information (article title/ journal title/ author/ publication information)
 - Subject descriptors
 - Keywords (author supplied?)
 - Abstracts (author supplied?)
 - Full-text

Topical Content Strength

- Do you feel there is a particular content area that you feel the service covers especially well or leans heavily toward (e.g., Humanities, Social Sciences, Sciences).
- Do you feel there is a particular content type that you feel the service covers very well or leans heavily toward (scholarly journal contents, mainstream journal contents, newspapers, conference proceedings).
- What subject/ content areas, if any, do you feel the service may be somewhat weak? Are there current efforts to mitigate these weaknesses (e.g., future publisher agreements on the horizon)?

Google Books / Google Scholar

- Do any agreements exist at this time to harvest the data associated with the Google Books or Google Scholar projects into your central index? If so, could you please describe the level of indexing (e.g., basic citation level metadata – title, author, publication date, abstracts, and full-text).

WorldCat Catalog

Does your service include the OCLC WorldCat catalog records?

If so, what level of information is included?

The complete record?

Holdings information?

E-Book Vendors

- Does your service include items from major e-book vendors?

Open Access Content

- **Open Access Content Sources.** Does your service automatically include (out of the box, no additional charge) materials from open access repositories? If so, could you please list some of the major repositories included (e.g., arXiv E-prints, IndiaRxiv, Directory of Open Access Journals, Hathi Trust Materials, Hindawi Publishing Corporation, etc.).
- **Open Access Content Sources: Future Plans.** In addition to the current open access repositories that may be included in your service, are there other repositories whose content you are planning to incorporate in the future?
- **Exposure to other Libraries' Bibliographic/ Digital Collection/ IR Content.** Are ILS bibliographic records from other customers using your discovery platform exposed for discoverability in the searchable discovery instance of another customer? Are digital collection records? Institutional repository records?

Relevancy Ranking

- **Relevancy Determination.** Please describe some of the factors which comprise the determination of relevancy within your service. What elements play a role, and how heavily are they weighted for purposes of determining relevancy?
- **Currency.** Please comment on how heavily currency of an item plays in relevancy determination. Does currency weigh more heavily for certain content types (e.g., newspapers)?
- **Local Library Influence.** Does the local library have any influence or level of control over the relevancy algorithm? Can they choose to “bump up” particular items for a search?
- **Local Collection Visibility.** Could you please offer some comments on how local contents (e.g. ILS bibliographic records; digital collections) remain visible and discoverable within the larger pool of contents indexed by your service? For example, local contents may measure a million items, and your centralized index may cover half a billion items.

Exposure of Items with Minimal Metadata

- Some items likely have lesser metadata than other items. Could you please offer some comments on how your system ensures discoverability for items with lesser or minimal metadata?

Full Text Searching

- Does your service offer the capability for the user to search the full-text of materials in your service, i.e., are they searching a full-text keyword index? If so, approximately what percentage of items within your service are “deep indexed?”
- Please describe how your system deals when no hits are retrieved for a search. Does your system enable “best-match” retrieval – that is, something will always be returned or recommended? What elements play into this determination; how is the user prevented from having a completely “dead-end” search?

Authentication and Rights Management

- Does your system offer an unauthenticated view/ access? Please describe and offer some comments on what materials will not be discoverable/ visible for an unauthenticated user.
- Licensed Full-Text
- Records specifically or solely sourced from Abstracting and Indexing Databases
- Full citation information (e.g., an unauthenticated user may see just a title, an authenticated user would see fuller citation information)
- Enrichment information (such as, book image covers, table of contents, abstracts, etc.)

Authentication and Rights Management

- Please discuss how rights management is initialized and maintained in your system, for purposes of determining whether a local library user should have access to the full-text (or otherwise “full resolution” if a library doesn’t license the full-text, such as, resolution to a detailed citation/ abstract).

Miscellaneous

The following feature/attribute is important to have in the discovery system .

- The vendor has an existing mobile version of their discovery tool for use by smartphones or other small internet-enabled devices.
- The vendor has designed the product such that it can be incorporated into other sites used by students, such as, WebCampus and/ or social networking sites. Such ‘designs’ may include the use of persistent URLs to embed hyperlinks, the ability to place the search box in another website, or specifically designed widgets developed by the vendor.
- Indexing and availability of newly published items occur within a matter of days as opposed to a week or perhaps a month.

STATISTICS

- The following statistic is important to have for the discovery platform [Strongly Disagree/ Disagree/ Neither Agree or Disagree/ Agree/ Strongly Agree].
- Number of searches, by customizable timeframe.
- Number of items or article level records accessed (that is, a user clicks on something in the returned list of results).
- Number of searches generating 0 results. *Investigations into Library Web-scale Discovery Services | Vaughan 56.*
- Number of items accessed by type.
- Number of items accessed by provider of contents (that is, number of articles from particular database/full-text vendor).

Thank you very much for your Attention