One Box to Search Them All:
Implementing Federated Search at an Academic Library [Post Print Version]

Abstract

Purpose: In May, 2008, the Ad Hoc Committee on Federated Search was formed to prepare a preliminary report on federated searching for a special meeting of Librarians Academic Council at Memorial University Libraries. The primary purpose was to discuss current implementation of federated searching at this institution, explore what other institutions have done, examine federated search technologies, and offer recommendations for the future of this resource.

Design/methodology/approach: Information was drawn from a recent usability study, an informal survey was created, and a literature/technology review was conducted.

Findings: These four recommendations were proposed and unanimously accepted: 1) actively develop our current federated search implementation by developing a web presence supporting 'federated search in context', 2) reevaluating the need for consortial purchase of a federated search tool, 3) continuing to assess the current federated search marketplace with an eye to choosing a nextgeneration federated search tool that includes effective de-duping, sorting, relevancy, clustering and faceting, and 4) that the selection, testing, and implementation of such a tool involve broad participation from the Memorial University Libraries system.

Originality/value: Provided is an inside look at one institution's experience with implementing a federated search tool. This paper should be of interest to anyone working in academic libraries, particularly the areas of administration, public services, and systems.

Introduction

Federated search at Memorial University Libraries has had many bumps. After a few years of low usage and concerns voiced by librarians, the Librarians Academic Council formed an ad hoc committee to assess our current implementation and recommend future directions for federated searching at Memorial University. This paper, an abridged version of the report submitted to Council, will describe the current implementation, outline what has been done elsewhere, and examine federated search technologies.

Part A: Implementation at Memorial

Memorial's current implementation of SirsiDynix Single Search was purchased through a consortial agreement without broad in-house consultation. Librarians had a mixed reaction to the product when it was first launched in Spring 2006, and reservations were expressed about federated search in general.

Consequently, the resource was not heavily used, nor was it consistently promoted to patrons. Frequently described as a resource discovery tool, SingleSearch is a single interface allowing users to simultaneously search multiple resources. A search can be queried two ways. Firstly, a user can search by selecting any number of subject areas or groupings (i.e. earth sciences, library catalogues, all subjects). Secondly, a user can search by selecting individual resources (i.e. America: History and Life, ebrary, Memorial University Libraries Catalogue). Memorial's federated searching tool is accessible through a quick search menu present throughout the library website. More recently, SingleSearch has been incorporated into Memorial's 'Explore a Topic' implementation of SirsiDynix's Rooms content management software. By default, search results are displayed according to a fastest first configuration, meaning that results which are returned the quickest appear first on the results list.

Is it Usable?

Usability is defined as "the ease with which a computer interface can be efficiently and effectively used, especially by a novice. The first priority in designing for usability is to provide clear, consistent navigation of content" (Reitz, 2007). With this in mind, a usability study was conducted in March 2006 to evaluate the use and perception of Memorial's federated search tool based on its effectiveness, efficiency, and satisfaction. To investigate these aspects, participants were asked to complete these basic SingleSearch tasks:

i) Selecting subject groupings and/or individual resources based on a provided topic.
ii) Conducting a search also based on a provided topic.
iii) Identifying one book and one article from the first page of results.
iv) Checking local holdings for selected book and article.
v) Using the 'Get it @ Memorial' Resolver[1] button to determine location of an item.

(Byrne and McGillis, 2008). Drawing on interconnected tasks, the study was able to present participants with realistic, problem-based research scenarios.

Usability Observations

This usability study resulted in several surprising and perhaps not so surprising findings which offer useful insight. Of these, consistent trends included: scale, lots of clicking, determining local holdings, easy access, and lack of sophisticated search strategies.

i) Scale. Participants were generally successful in selecting appropriate categories, with few opting to choose individual resources. This suggests that these library users are attracted to searching broader categories as compared to information seeking at a micro level. For example, the depth of categories was sometimes so vast that 25 resources were being searched at once (Byrne and McGillis, 2008).

ii) Lots of clicking. Byrne and McGillis observed that participants required an average of 5 minutes to find an article or give up, and 4 minutes for books (2008). It was also noted that "the high number of clicks and multiple attempts indicate extreme difficulty finding holdings information and even a confusion as to what holdings were" (2008).

iii) Determining local holdings. In addition to excessive clicking, difficulty was exhibited in

determining local holdings, and only one-fifth of users correctly interpreted catalogue book holdings (Byrne and McGillis, 2008). This suggests a general need for increased information literacy skills, and is perhaps not necessarily a direct reflection of SingleSearch.

iv) Easy access. Participants were much more successful locating items in the Resolver as compared to the Catalogue, and were very likely to ditch the article if full-text could not easily be found (Byrne and McGillis, 2008).

v) Lack of sophisticated search strategies. All participants entered the topic exactly as written meaning that no sophisticated search strategies or Boolean operators were used (Byrne and McGillis, 2008). Participants searched on a purely literal level and this raises general questions about information literacy, likely beyond the federated searching realm.

Potential Directions

Observations of this study provide several potential directions for federated searching at Memorial, with a few highlighting conclusions of this study being simplifying resource selection, offering an efficient path to the article, value in including format identifiers, and response time.

i. Simplifying resource selection. The central feature of federated searching is the ability to search multiple resources simultaneously. At the same time, searching within a subject specific context has many advantages and Explore a Topic offers major potential in this area. Worth noting is that a recent usability study discovered that the implementation of SingleSearch into this environment has improved article searching from five minutes to three (Byrne and McGillis, 2008). This suggests the benefit of searching a more relevant group of resources as compared to searching any and all potentially useful sources.

ii. Offering an efficient path to the article. The majority of problems were encountered once users left the SingleSearch interface, and a follow-up focus group was since been conducted to investigate usability in this area and resulted in replacing the original Get it @ Memorial plain text with an image (Byrne and McGillis, 2008).

iii. Value in including format identifiers. Users had little difficulty identifying formats which is likely due to the presence of format tags in each record (Byrne and McGillis, 2008). Continued inclusion of this feature would likely contribute to positive resource discovery experiences.

iv. Response time. Interestingly enough, load speed was not mentioned in the post-test even though it appeared to be an issue to study designers based on recorded screen activity (Byrne and McGillis, 2008).

II. What Service Providers Think

To gain a balanced look at this topic, an informal survey was made available to all librarians and library assistants at Memorial University Libraries in May 2008. Intended to gain current insight into how service providers view this tool, this survey asked five questions and a total of thirteen respondents participated with a 100% completion rate. While findings were reminiscent of comments often exchanged through general discussion, much thought provoking feedback was provided.

When asked for feedback about the most desired SingleSearch traits, consistent reference was made to the ability to search multiple sources simultaneously, use of subject groupings, and a friendly interface. One comment drew attention to the ability to add a SingleSearch search box to subject specific environments (i.e. subject specific web pages). This 'like' echoed the need to simplify resource selection as expressed in Byrne and McGillis' usability study. This is not to say, however, that SingleSearch pet peeves do not exist, and when asked for feedback about less than stellar traits, feedback consistently drew attention to too many results, vast resource selection, slow speed, loss of native search abilities, and lack of de-duping, relevancy ranking, and clustering. One comment that stood out was: "it just seems like it should be able to do so much more than it actually does". It is revealing that concern with slow speed was a service provider issue, yet was not mentioned by participants of the SingleSearch usability study (Byrne and McGillis, 2008). This informal survey also asked service providers whether they think and want federated searching to have an increasingly prominent role at this institution. Almost all responses supported both desires, suggesting that federated searching will play a progressively important role in Memorial's future, as the desire to make it so already exists.

What does all of this mean? The reality is that service providers promote and encourage use of resources that they consider to be valuable, yet if a product is viewed as problematic to work with, it will likely suffer from low usage or be phased out. Based on survey feedback, there is a distinct sense that federated searching is unlikely to disappear, the questions that remains is determining what product will take us to a usable future the fastest. Until then, it will be of benefit to implement the SingleSearch search box at a subject specific level to provide a more relevant searching experience.

Part B: What Others Have Done

I. Selecting a Product

The Committee

At most institutions, a committee is formed to implement a federated search solution, and should include reference librarians, subject specialists and systems people. It is also crucial that the committee consult extensively with all stakeholders (Boss and Nelson, 2005; Caswell and Wynstra, 2007; Marshall, Herman and Rajan, 2006; Avery, Ward and Janicke Hinchliffe, 2007; Elliott, 2004).

The Specification

The committee must first determine what the organization needs the product to do. In most implementations, the most important requirement is that the bulk of available electronic resources be searchable. This requirement is universal regardless of the size of the organization (See Hill, 2007; McHale, 2007; Boock, Nichols and Kristick, 2006). A Googlesque search box is another common requirement (Grimes, 2007; Herrera, 2007; Boock, Nichols and Kristick, 2006; Cervone, 2005). Resources to be searched this way are usually determined by response times during testing. Advanced search functionality allowing the user to pick individual resources or a preselected subject grouping is also required (Grimes, 2007; Herrera, 2007; Newton and Silberger, 2007; Boock, Nichols and Kristick, 2006).

It may be necessary to specify what customization options are desired. For example, a key requirement for the Intel Library was for the search tool to be seamlessly integrated into their existing intranet portal (Hill, 2007). To achieve the desired level of customization, some vendors may require institutions t o buy extra products, this was the case for California State University – San Marcos which purchased Ex Libris' MetaLib X-Server to customize their MetaLib implementation for broader use across different systems (Walker, 2007). Cost can vary greatly between products or even between packages for the same product, and total cost of ownership is an important consideration. Maintaining a product carries significant costs, especially if it is hosted locally, and while a hosted service may be less costly, it may be less customizable (Caswell and Wynstra, 2007).

Trials

After specifications have been determined, it is crucial that the library trial any product which seems to meet this criteria. Some vendors have little trouble making a customized trial that includes all of a library's resources (Hollandsworth and Foy, 2007; McHale, 2007; Scherlen, 2006). The trial will reveal some of the issues with the software and some of the customization that will be necessary.

Picking a Product

Product selection should take place only after careful deliberation of the attributes of both the product and vendor (Elliott, 2004).

Consortial Purchase

The limited reports of consortial implementations paint a mixed picture, and in general, there appears to be little incentive to implement federated search as a consortium. The Georgia Virtual Library, a consortium of 400 school, public and academic libraries, attempted to implement a single federated search solution across all libraries. Unable to balance all institutional needs in one product, two products were selected: one immediately implemented for school and public libraries and one for academic libraries that as of this writing appeared to be in limited testing (Fancher, 2007). More disappointingly, the Five College Libraries of Western Massachussetts federated search implementation ended in failure. The consortium could not sufficiently customize the product to the standards of each member, and while some testing occurred, the product was set aside by most members (Mestre et al., 2007).

II. Implementation

"Those libraries that have been willing to report on their implementation of federated searching applications have described missed deadlines, soft launches and compromises made along the way" (Warren, 2007, p. 258)

Warren is more pessimistic than most authors, and stresses that the implementation phase takes a long time. At a minimum, libraries should plan for at least one librarian to work on the implementation for 6-12 months(Elliott, 2004). Regardless of resources devoted to the project, it is essential to have an implementation plan and to stick with that plan as closely as possible (Caswell and Wynstra, 2007; Hill, 2007; Marshall, Herman and Rajan, 2006; Avery, Ward and Janicke Hinchliffe, 2007; Boock, Nichols and

Kristick, 2006).

Deciding on Resources to Include

Resource selection is a delicate balance between speed and comprehensiveness, and as more resources are searched, result retrieval takes more time. Some vendors limit how many resources can be simultaneously searched to try to improve speed (Walker, 2007). For resources that can't be searched through the federated search product, the library either has to write a custom connector or pay the vendor to write one. Most vendor hosted products force libraries to pay if they want to include locally hosted content (Jung et al., 2008; Boock, Nichols and Kristick, 2006). A Googlesque search of a few full text resources is usually provided, and resources are typically selected due to response time and coverage (Grimes, 2007; Herrera, 2007; Jung et al., 2008; Newton and Silberger, 2007; See Boock, Nichols and Kristick, 2006; Cervone, 2005).

The organization of the advanced search feature is more controversial. Most libraries provide a general list of subjects assigned certain databases by subject specialists. Creating subject lists is not without problems, especially in products that limit the number of resources that can be searched simultaneously.

> Users don't inherently know which databases to search when faced with having to search each individually. Although metasearch systems allow users to search multiple databases simultaneously, that does little in itself to resolve the question of why a user would choose to search, say, Project Muse and JSTOR simultaneously over WorldCat and Academic Search Elite simultaneously (Walker, 2007, pp. 328-329).

Warren (2007) argues that to offer adequate subject coverage, long lists must be created for users to pick their subjects with sufficient specificity, which would recreate a problem federated search was intended to solve. Subject lists are a compromise most libraries are willing to make, and there is some reluctance at some institutions to let users pick individual databases (See Jung et al., 2008; Elliott, 2004; Cervone, 2005).

Testing

Avery (2007) suggests that three kinds of testing should occur on the system before it goes live. Firstly, testing should verify that the system has all the features the vendor said it did. Secondly, testing should focus on functionality of the product. Are searches being interpreted properly by the native databases? Is the search returning the expected results? Are the results being sorted? Are duplicates being removed? (Avery, Ward and Janicke Hinchliffe, 2007) Thirdly, usability testing should take place. Elliott (2004) was very critical of the usability testing occurring at the time of her report as it was not being done early enough in the implementation cycle, did not involve enough different users, and results were not being followed up on. Her summary of the situation was that "[w]hile libraries are indeed conducting usability testing and other types of user studies, it is clear that librarian preferences rather than user preferences still rule the design of library-oriented metasearch tools"(Elliott, 2004, p. 23). The literature shows that extensive usability testing is performed in most cases (George, 2008; Avery, Ward and Janicke Hinchliffe, 2007; Boock, Nichols and Kristick, 2006). Avery (2007) reports spending 15 weeks on just usability testing and tweaking the interface.

Marketing

Before going live, it is important to market the product as widely as possible. Most implementations used campus media and flyers; others opted for extra promotional materials like tent cards, pens, mouse pads, etc. to place around campus and for use in promotional events (Cox, 2007; Wisniewski, 2007). The service name should be carefully selected to be memorable but not misleading. The University of Pittsburgh branded their search "ZOOM!" which was easy for patrons to remember but usability testing revealed it gave patrons the mistaken impression that the service was very fast (Wisniewski, 2007).

III. Feedback

Disappointed Librarians

For the most part, librarians have been disappointed with federated searching. It is not uncommon for librarians to expect federated search products to have the same functionality as the respective native interface (See Baer, 2004; Boss and Nelson, 2005; Fahey, 2007; Tang, Hsieh-Yee and Zhang, 2007; Avrahami et al., 2006). This demonstrates that librarians do not understand, at a fundamental level, how federated searching works behind the scenes. Tang (2007) found that librarians tended to use federated searching as a tool of last resort whereas other users were happy to use it as their primary means of searching. Others have claimed that federated searching is nothing more than pandering to the lowest common denominator (Marshall, Herman and Rajan, 2006; Warren, 2007). Warren (2007) contends that Google and some native database interfaces offer many of the options that federated search should accommodate such as automatic stemming of search terms, suggested spellings, proper relevance ranking and context sensitive help. Search speeds are another issue. Most authors have noted that even searching a small number of resources produces long wait times and that search times increased dramatically as more resources were searched (Boss and Nelson, 2005; Calhoun, 2005; Fahey, 2007; Foust, Bergen and Maxeiner, 2007; Herrera, 2007; Mestre et al., 2007; Boock, Nichols and Kristick, 2006; Scherlen, 2006).

From a technical perspective, many librarians involved with maintaining these systems complain about a lack of standardization between vendors in how searches are transmitted and how results are received (Bracke, Howse and Keim, 2008; Calhoun, 2005; Jung et al., 2008; McHale, 2007; Elliott, 2004; Warren, 2007; Boock, Nichols and Kristick, 2006; Webster, 2007). For resources not compatible with federated search, the library can either use HTML 'screen-scrape' or omit these resources altogether. Both of these options frustrate librarians. If the library chooses to screen-scrape, the connectors constantly break and have to be changed whenever the native interface changes (Hollandsworth and Foy, 2007; Marshall, Herman and Rajan, 2006). If the library opts to not include a resource, there is a chance that users will assume that it is not important and therefore not use it (Baer, 2004; Cervone, 2005). Librarians that seemed happiest with their federated search products tended to see the limitations of the technology and viewed it as a complement, not a replacement, to existing database interfaces.

Ambivalent Users

"The irony remains that a tool designed to ease the burden of searching can be itself difficult to use."(Elliott, 2004, p. 13)

Usability testing has shown that user problems with federated search products fall into three categories: interface, search, and a lack of knowledge. Severe interface problems have emerged through usability studies. A general problem is that users rarely see the navigation elements or advanced features and when they do see them they don't understand the symbols (George, 2008; Mestre et al., 2007; Ponsford and vanDuinkerken, 2007; Elliott, 2004; Wrubel and Schmidt, 2007). Using the web browser's navigation can often cause more problems. Another serious interface issue is the display of results. Sufficient information is often unavailable for a patron to determine whether the result is a book, journal article, or some other type of item in a collection (Ponsford and vanDuinkerken, 2007; Boock, Nichols and Kristick, 2006; Walker, 2007; Wrubel and Schmidt, 2007). Also, most products offer no way to refine a search once the results screen has come up short of redoing the search (Boock, Nichols and Kristick, 2006; Walker, 2007; Wrubel and Schmidt, 2007). The search product often doesn't execute the search in the manner that users expected it to. This problem arises because users assume the single search box will work like Google (Hill, 2007; Ponsford and vanDuinkerken, 2007; Sadeh, 2007; Walker, 2007). How the search is actually executed varies greatly by product. As if these problems were not severe enough, most patrons don't understand what they are searching (Newton and Silberger, 2007; Ochoa et al., 2007). For those patrons that are interested in what they search, it is important that libraries make it clear what is being searched.

IV. Examples of Federated Search Systems

There are several interesting uses of federated search in the literature. Arizona Health Library System developed a tool that searches all their electronic book collections and sorts results based on the evidence pyramid (Bracke, Howse and Keim, 2008). Rainwater (2007) describes how Brown University created a workflow for tracking and organizing e-resources and ultimately integrating them into their federated search. One interesting case study is Oregon State University. In 2003, they purchased a solution from a vendor, implemented it, and tried to make it work. The search speed was very slow and though a number remedies were tried nothing seemed to fix it. Finally, the library had enough and in 2006 built their own tool called LibraryFind which is still under development but in production. To improve performance, the library has started mounting some indexes locally, and initial reports have been very positive. One of the major benefits of making their own product is that local repositories can be included for no extra cost (Jung et al., 2008; Boock, Nichols and Kristick, 2006).

Part C: The Technology

Federated Search Technologies can be divided into two major categories: cross search, which searches distributed sources simultaneously and presents the results in a common results interface; and harvested search, which retrieves the contents of multiple distributed databases, normalizes the records, and stores them in a large union index against which the searches are then run. There are benefits and drawbacks to each approach.

I. Cross search

Cross search engines are those which search distributed targets on the fly, and return the results to a common search interface. Cross-search is also known as cross-database searching, parallel searching, and broadcast searching.

Connectors

Each target database requires a "connector". The connector tells the search engine how to request results from a given source, and how to interpret those results for display. There are three broad types of connectors:

- XML Gateways – Increasingly available in large commercial databases, XML gateways are fairly stable and return results quickly. Major standards for XML Gateways include SRU/SRW, OpenSearch, and MetaSearch XML Gateway (MXG).

- Z39.50 – Library catalogues and other library-specific technologies often expose themselves to federated search engines through a Z39.50 server. These search targets tend to be quite stable, and the connectors rarely require updating.

- Screen Scraping – Used when the target database does not support any of the other search protocols. Relevant elements are parsed out of the HTML code underlying the native interface. Connectors that use screen scraping are very unstable, and have to be adjusted with every modification of the target database interface.

Presenting Results in Cross-Search

Cross search engines receive results from a number of different sources. There are various ways of presenting these results.

Fastest first: Because some databases return results more slowly than others, a fastest first configuration will make sure that your user sees some result within a reasonable timeframe. The drawback to fastest first is that the fastest database will not necessarily be the most relevant source.

Relevancy ranking: Federated search has very limited information with which to perform its ranking. The engines must determine relevancy using only the words that appear in citation fields like document title, journal title, or abstract. Often, the search word doesn't even appear in the citation. Native sources have access to the full text of their articles, so they can rank much more precisely. Some federated search systems don't perform ranking at all, but simply return documents as ranked by the source.

De-Duping: For federated search engines, true de-duplication is virtually impossible. In order to dedupe, the engine would have to download all search results and compare them. Because databases return results 10 or 20 records at a time, completing a true de-dupe operation on 50, 000 hits would take hours. Cross-search vendors usually just de-dupe the first result set returned by the search.

Clustering: Clustering algorithms look for similar words and phrases in the citations returned by a search, and attempt to group documents that have many words and phrases in common. Clustering software was originally developed to operate against full text documents, but federated search engines try to create relevant clusters out of the smaller number of words available in citations. Relevance ranking and clustering are not mutually exclusive features, both the hit list and the clusters should be ordered according to relevance.

Faceting: Faceted navigation organizes results according to common metadata attributes like author, publication date, journal, or other citation elements. It provides a powerful discovery tool, allowing users to drill-down into the results and focus a query more precisely. Faceted navigation works best against highly structured data, so is well suited for cataloguing records and citations. Faceting in the federated search environment is subject to the usual limitations: the sparseness of data in a citation, and the small number of citations returned at one time.

II. Harvested/Union Indexes

The second major approach to federated search is to harvest all of the relevant sources of data, normalize them into a single metadata schema, and index all of them together in one large union index. This approach offers huge advantages in speed and in the logic that can be applied to the presentation and sorting of results.

In most cases a harvested/union index solution will require the provider to download the metadata records at a minimum, and ideally the full text documents too. There is a fair amount of expense involved in maintaining the hardware, software, and network infrastructure to support the frequent harvesting of large record and document sets from many sources. Each data source requires a unique parsing routine to extract and normalize the data, and indexes must be constantly updated and optimized.

More challenging than the technical obstacles are the legal aspects of data harvesting, particularly rights management. It would be very difficult for a single library to negotiate the right to harvest data from each of its licensed databases providers, particularly full-text data. Federated search solutions based on this model tend to be developed by commercial vendors or large library consortia.

Harvesting Protocols

There are several major harvesting standards currently in play:

- OAI-PMH – Commonly implemented in digital repositories, Open Archives Initiative Protocol for Metadata Harvesting allows a service provider to send an http request to a data provider, who returns an XML–encoded data stream containing the metadata for a specific collection of objects or articles.

- METS- Similar to OAI-PMH in purpose and function, METS supports XML-encoded metadata harvesting, but unlike OAI-PMH, METS can harvest both metadata and object.

- LOCKSS – LOCKSS collects content by crawling a web site and downloading each page it finds there. The data provider must implement a "manifest" page for each collection, explicitly granting permission for LOCKSS crawlers to visit. The owner of the LOCKSS box maintains "plugins" that instruct the LOCKSS software how to crawl and audit content from each provider.

- Custom Formats – A vendor might prefer to supply metadata or content in an agreed-upon custom XML or delimited text format.

Presenting Results from Harvested/Union Indexes

Speed: A single index will return results much more quickly than multiple disparate indexes, particularly if the index is hosted on a local server so that Internet latency is not a factor.

De-duping: True de-duping is possible in this environment, because the index already contains all of the results sets. It can compare and de-dupe them very quickly, as it does not have to wait for the results to be returned from each source in sets of 20 hits at a time.

Relevancy: Results can be relevancy ranked with much more granularity and accuracy if they are stored in a single large index, particularly if that index has access to the full text of structured documents. The organization maintaining the index will also have access to tweak the ranking algorithms and customize the way in which results are returned.

Clustering and Faceting: Both of these features are much easier to implement against a single large data store, because all of the necessary information has already been parsed out, normalized, and indexed appropriately. In a distributed (cross-search) solution the data has to be normalized and indexed on the fly. A full text data store will improve the engine's ability to assess similarities, although citation information can also be used.

Web Search Engines

Web search engines like Google and Yahoo are some of the most familiar examples of large union indexes. These services send out many "spiders" or "web crawlers", small programs that traverse the web by following links. The spiders send full text information and HTML structure back to huge union indexes that underpin the major web search engines. Web search engines do a good job of indexing large numbers of unstructured HTML and PDF documents, but the algorithms that they use are far less effective at indexing and ranking enterprise data (e.g. all of the sources of structured data that form part of an organization's intranet, or in the case of libraries all of the structured metadata records to which a library has access through ownership or subscription).

Most web search engine companies provide "local" versions of their programs that can be customized to search a select set of targets. Google Co-op and Yahoo Alpha will both allow the end-user to customize the set of resources that they wish to search, and also to personalize the experience by changing the default position of search sources, etc. These services are insufficient for library use however, as they do not handle authentication to restricted sources, and are largely incapable of mining the deep web where most of our indexed content resides.

III. Federated Search Marketplace

Commercial Cross-Search Engines

Most complete federated search solutions support multiple search protocols. Typically they offer integrated OpenURL resolution, spell checking, saved searches, alerts, de-duping, and single click access to the native interface. Many recent offerings incorporate clustering and faceting for improved research discovery and guided search services. Some interfaces also support tagging and other 2.0 features. Some

provide support for harvesting and locally indexing results from some sources while cross-searching others. It is not uncommon for a federated search solution to use one company for the creation and maintenance of connectors (e.g. MuseGlobal), and a second vendor to provide the discovery layer interface (e.g. Vivisimo). The following is a selected list of some of the major players in cross-search:

MuseGlobal Database Connectors (http://www.museglobal.com/solutions/index.html)
Implementations: Sirsi Single Search, Endeavor, Innovative, Ovid Search Solver, Swetswise Searcher
MuseGlobal is a company that primarily builds and maintains a large a database of search connectors that are licensed to other companies for use behind their own discovery interfaces.

Vivisimo Velocity Search Platform (http://vivisimo.com/products/products)
Implementations: Serials Solutions, ExLibris Metalib, SwetsWise Searcher NLM, www.usa.gov
Vivisimo is a company best known for enterprise search interfaces, but its clustering technology has been adopted as a discovery interface for a number of federated search products.

Serials Solutions 360 Search (http://www.serialssolutions.com/ss_360_search.html)
Implementations: University of Arizona, University of Wyoming, Central Missouri State
Serials Solutions develops and maintains its own bank of search connectors. The 360 interface is a remotely-hosted product built on Vivisimo clustering technology.

Serials Solutions Webfeat (http://www.webfeat.org/)
Implementations: Cambridge University, University of Calgary
Originally owned by Data Associates, Webfeat was acquired by Serials Solutions (Proquest) in 2008. Webfeat has its own extensive collection of connectors that will likely be merged with the 360 connector library and centrally maintained by Serials Solutions.

NOTE: Serials Solutions has announced an intention to support both 360 Search and Webfeat through 2008, but to integrate the best aspects of each into a single product in 2009.

SirsiDynix Single Search (http://www.sirsidynix.com/Solutions/Products/portalsearch.php)
Implementations: Atlantic Scholarly Information Network, University of Alberta
SirsiDynix Single Search relies on the extensive MuseGlobal connector library.

ExLibris Metalib (http://www.exlibrisgroup.com/category/MetaLibOverview)
Implementations: Oxford, Texas AandM, Northwestern, Brown
Metalib's discovery interface incorporates the Vivisimo clustering engine.

Deep Web Technologies Explorit Research Accelerator (http://www.deepwebtech.com/)
Implementations: Stanford, Scitopia.org, Science.gov, WorldWideScience.org
Deep Web Technologies, best known for creating single search solutions for government agencies, have recently partnered with Stanford to create a federated search tool for the academic environment.

All Access Connector
(http://www.museglobal.com/news/news.php?content=2008_innews/20080505.html)
MuseGlobal and Adhere Solutions have partnered to integrate Muse's 5400+ prebuilt connectors into the Google Search Appliance. Announced May 2008.

OpenTranslators (http://www.librarytechnology.org/ltg-displayarticle.pl?RC=12980)

Implementations: Ohio Public Library
The OpenTranslators project provides a gateway to access the library of WebFeat connectors. The translators are neither open source nor free, but the access layer is open source. OpenTranslators allow libraries to use Webfeat connectors with a federated search interface of their choice.

Orgeon State LibraryFind (http://search.library.oregonstate.edu/record/search)
Implementations: Oregon State
Launched in 2007 with a grant from Oregon State University Library, LibraryFind is an open source federated search offering. The number of available connectors is limited compared to the large commercial offerings, although libraries can develop and add as many connectors as they wish.

Harvested Search Solutions

Because of the technical and legal obstacles to establishing a large full text repository of licensed content, very few individual libraries have tried to create federated search solutions that rely on the maintenance of large union indexes. There are however a few commercial and consortial players who are currently engaged in the creation of such services. Examples include:

Google Scholar (Free Indexing)
Unlike most of the services mentioned in this section, Google Scholar is not a repository, but is simply a large index created by web crawlers that are restricted to sources broadly defined as "scholarly".

OhioLINK (Consortial Harvesting of Licensed Content)
OhioLINK maintains an Electronic Journal Center (EJC), which contains more than 6,400 scholarly journal titles from more than 80 publishers across a wide range of disciplines. OhioLINK has declared its intention to maintain the EJC content as a permanent archive and has acquired perpetual archival rights in its licenses from almost all publishers.

Ontario Council of University Libraries (Consortial Harvesting of Licensed Content)
Scholar's Portal includes 7,500 e-journals from about 20 publishers, and metadata for the content of an additional three publishers. The primary purpose of the portal is access, but OCUL has made an explicit commitment to the long-term preservation of the e-journal content it loads locally.

PubMed Central (Free Indexing and Full-text Access)
PubMed Central is the NIH's free digital archive of biomedical and life sciences journal literature, currently encompassing approximately 220 titles from 40 publishers.

CSA Illumina (Fee-based Indexing and Full-text Access)
CSA Illumina provides access to more than 100 databases published by CSA and its publishing partners.

OCLC FirstSearch Databases (Fee-based Indexing and Full-Text Access)
OCLC's Electronic Collections Online (ECO) is an electronic journals service that offers single search access to a collection of more than 5,000 titles in a wide range of subject areas, from over 70 publishers. Index Data is currently helping OCLC incorporate metasearch into WorldCat Local to provide access to databases that are not indexed in WorldCat.org.

Conclusion

In light of the above information, Memorial University Libraries have resolved to do the following:

1. More actively develop our federated search implementation by redeveloping our web presence to support "federated searching in context".

2. Continue to assess the current federated search marketplace with an eye to selecting a next generation federated search tool that includes effective de-duping, sorting, relevancy ranking, clustering and faceting of search results.

3. Reevaluate our current consortial arrangements going forward.

4. Commit to involving a broad range of people in the selection, testing and implementation of any new products.

Following the upcoming redesign of the library website to better incorporate SingleSearch, implementation of this tool will be reevaluated in December, 2008. At that time, a formal decision will be made as to the future of federated search at Memorial University Libraries.

Notes

[1] A link resolver provides "context-sensitive linking between a citation in a bibliographic database and the electronic full text of the resource cited (article, essay, conference paper, book, etc.) in an aggregator database or online from the publisher, taking into account which materials the user is authorized by subscription or licensing agreement to access." (Reitz, 2007)

References

Avery, S., Ward, D. and Janicke Hinchliffe, L. 2007, "Planning and Implementing a Federated Searching System: An Examination of the Crucial Roles of Technical, Functional, and Usability Testing", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 179-194.

Avrahami, T.T., Yau, L., Si, L. and Callen, J. 2006, "The FedLemur Project: Federated Search in the Real World", Journal of the American Society for Information Science and Technology, vol. 57, no. 3, pp. 347-358.

Baer, W. 2004, "Federated searching: Friend or foe?", College and Research Libraries News, vol. 65, no. 9, pp. 518-519.

Boock, M., Nichols, J. and Kristick, L. 2006, "Continuing the Quest for the Quick Search Holy Grail: Oregon State University Libraries' Federated Search Implementation", Internet Reference Services Quarterly, vol. 11, no. 4, pp. 139-153.

Boss, S.C. and Nelson, M.L. 2005, "Federated Search Tools: The Next Step in the Quest for One-Stop-

Shopping", Reference Librarian, vol. 44, no. 91, pp. 139-160.

Bracke, P.J., Howse, D.K. and Keim, S.M. 2008, "Evidence-based Medicine Search: a customizable federated search engine", Journal of the Medical Library Association, vol. 96, no. 2, pp. 108-113.

Byrne, G. and McGillis, L. 2008, Single Search Usability Report, Internal document, Memorial University of Newfoundland, St. John's, NL.

Calhoun, K. 2005, "An integrated framework for discovering digital library collections", Journal of Zhejiang University SCIENCE, vol. 6A, no. 11, pp. 1318-1326.

Caswell, J.V. and Wynstra, J. 2007, "Developing the Right RFP for Selecting Your Federated Search Product: Lessons Learned and Tips from Recent Experience", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 49-71.

Cervone, F. 2005, "What We've Learned From Doing Usability Testing On OpenURL Resolvers and Federated Search Engines", Computers in Libraries, vol. 25, no. 9, pp. 10-14.

Cox, C. 2007, "Hitting the Spot: Marketing Federated Searching Tools to Students and Faculty", The Serials Librarian, vol. 53, no. 3, pp. 147-164.

Elliott, S.A. 2004, Metasearch and Usability: Toward a Seamless Interface to Library Resources., University of Alaska, Anchorage, AK.

Fahey, S. 2007, "F*****ED Searchers? The Debate about Federated Search Engines", Feliciter, vol. 53, no. 2, pp. 62-63.

Fancher, L. 2007, "Wanted, Dead or Alive: Federated Searching for a Statewide Virtual Library", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 133-158.

Foust, J.E., Bergen, P. and Maxeiner, G.L. 2007, "Improving e-book access via a library-developed fulltext search tool", Journal of the Medical Library Association, vol. 95, no. 1, pp. 40-45.

George, C.A. 2008, "Lessons learned: usability testing a federated search product", The Electronic Library, vol. 26, no. 1, pp. 5-20.

Grimes, M.F. 2007, "MSU Libraries' Implementation of Federated Search Software", Mississippi Libraries, vol. 71, no. 1, pp. 8-10.

Herrera, G. 2007, "Meta Searching and Beyond: Implementation Experiences and Advice from an Academic Library", Information Technology and Libraries, vol. 26, no. 2, pp. 44-52.

Hill, B. 2007, "Federated Search at the Intel Library", Information Outlook, vol. 11, no. 9, pp. 11-23. Hollandsworth, B.L. and Foy, J. 2007, "Griffin search: how Westminster College implemented WebFeat", Library Hi Tech, vol. 25, no. 2, pp. 211-219.

Jung, S., Herlocker, J.L., Webster, J., Mellinger, M. and Frumkin, J. 2008, "LibraryFind: System design and usability testing of academic metasearch system", Journal of the American Society for Information

Science and Technology, vol. 59, no. 3, pp. 375-389.

Marshall, P., Herman, S. and Rajan, S. 2006, "In Search of More Meaningful Search", Serials Review, vol. 32, no. 3, pp. 172-180.

McHale, N. 2007, "Accidental Federated Searching: Implementing Federated Searching in the Smaller Academic Library", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 93-110.

Mestre, L.S., Turner, C., Lang, B. and Morgan, B. 2007, "Do We Step Together, in the Same Direction, at the Same Time? How a Consortium Approached a Federated Search Implementation", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 111-132.

Newton, V.W. and Silberger, K. 2007, "Simplifying Complexity Through a Single Federated Search Box", Online (Weston, Conn.), vol. 31, no. 4, pp. 19-21.

Ochoa, M., Jesano, R., Nemmers, J.R., Newsom, C., O'Brien, M. and Victor Jr., P. 2007, "Testing the Federated Searching Waters: A Usability Study of MetaLib", Journal of Web Librarianship, vol. 1, no. 3, pp. 47-66.

Ponsford, B.C. and vanDuinkerken, W. 2007, "User Expectations in the Time of Google: Usability Testing of Federated Searching", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 159-178.

Rainwater, J. 2007, "Maintaining a Federated Search Service: Issues and Solutions", Internet Reference Services Quarterly, vol. 12, no. 3, pp. 309-323.

Reitz, J.M. 2007, November 19, 2007-last update, ODLIS: Online Dictionary for Library and Information Science [Homepage of Libraries Unlimited], [Online]. Available: http://lu.com/odlis/odlis_f.cfm [2008, 14/07].

Sadeh, T. 2007, "Transforming the Metasearch Concept into a Friendly User Experience", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 1-25.

Scherlen, A. 2006, "The One-Box Challenge: Providing a Federated Search That Benefits the Research Process", Serials Review, vol. 32, no. 4, pp. 247-254.

Tang, R., Hsieh-Yee, I. and Zhang, S. 2007, "User Perceptions of MetaLib Combined Search: An Investigation of How Users Make Sense of Federated Searching", Internet Reference Services Quarterly, vol. 12, no. 1, pp. 211-236.

Walker, D. 2007, "Building Custom Metasearch Interfaces and Services Using the MetaLib X-Server", Internet Reference Services Quarterly, vol. 12, no. 3, pp. 325-339.

Warren, D. 2007, "Lost in Translation: the Reality of Federated Searching", Australian Academic and Research Libraries, vol. 38, no. 4, pp. 258-269.

Webster, P.M. 2007, "Challenges for Federated Searching", Internet Reference Services Quarterly, vol. 12, no. 3, pp. 357-368.

Wisniewski, J. 2007, "Build It (and Customize and Market It) and They Will Come", Internet Reference Services Quarterly, vol. 12, no. 3, pp. 341-355.

Wrubel, L. and Schmidt, K. 2007, "Usability Testing of a Metasearch Interface: A Case Study", College and Research Libraries, vol. 68, no. 4, pp. 292-311.