

Federated Search Tools: The Next Step in the Quest for One-Stop-Shopping

Stephen C. Boss and Michael L. Nelson

SUMMARY. The emergence of federated search tools for the library market represents a significant step toward the longstanding goal of a common searching interface in the reference environment. These systems aim to provide integrated access to library resources of all types. This article describes and evaluates federated search tools from four vendors. It suggests criteria that may be useful for libraries attempting to compare and contrast this type of access tool.

KEYWORDS. Federated search tools, integrated searching, metasearching, library gateways, library portals, OpenURL resolvers, Z39.50

Stephen C. Boss is Systems Librarian at the University of Wyoming Libraries (E-mail: sboss@uwyo.edu). Michael L. Nelson is Social Sciences Reference and Collection Development Librarian at the University of Wyoming Libraries (E-mail: mnelson@uwyo.edu). The authors may be reached at the following address. University of Wyoming Libraries, Department 3334, 1000 E. University Avenue, Laramie, Wyoming 80271

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the assistance of the following: Nancy Chaffin, Metadata Librarian at Colorado State University Libraries, in providing information on MetaLib/SFX and hosting a site visit facilitating the authors' access to full functionality and all licensed databases at that site; Beth Forrest Warner, Director, Digital Library Initiatives and John S. Miller, Special Projects Librarian at the University of Kansas, in providing information on and access to their implementation of Endeavor's ENCompass product; Ellen Starkman, Acting Manager of Library Systems at the University of Illinois at Chicago; the State Library of New Jersey and our colleagues at the University of Wyoming Libraries who provided helpful feedback.

INTRODUCTION

Library reference staffs have always aimed to provide patrons with quality tools to find information and conduct research. Reference librarians participate in designing library websites to guide users to the OPAC, reference databases, in-house databases, and other sources. They recommend database purchases and try to offer as many online reference tools as possible to end-users. Yet all too often end-users are found staring at the screen in frustration, unsure where to begin. The problem typically is not a lack of access or resources, but rather the proliferation of choices. The competitive nature of the database market encourages vendors to differentiate themselves by emphasizing their unique features and user interfaces. Unfortunately, this leads to

confusingly complex arrays of “information silos” – disparate resources that do not interact and must be separately searched. Simply finding articles and books requires searching in at least two places, and frequently patrons must locate multiple sources that together meet their needs and then master their varying interfaces and data structures. All this can add up to a user interface that does not serve patrons well.¹

The proliferation of supercenters says much about what consumers want: singular places where they can shop for everything they need. While retailers have responded to meet this desire, the library world is just beginning to address how we can move toward a “one-stop-shopping” environment. Librarians need to take this path if they expect to remain a viable information source for the web-savvy student or member of the general public who has discovered the power and convenience of the general Internet. The debut of Google Scholar in late 2004, with its focus on library-oriented resources, only intensifies the need to better integrate library searching.

FEDERATED SEARCH TOOLS: A SOLUTION?

Happily, it appears that the next logical step in the continuing quest for a single common interface and truly integrated access to information in all its varied forms has been taken. Federated search tools independently search clusters of databases: the library’s OPAC, other library and consortia catalogs, indexing/abstracting databases, full-text, specialized search engines, digital objects, institutional repositories, document delivery services—pretty much anything that libraries own or access via a database. They search a variety of vendors

and systems with a single user interface, merge and deduplicate results, and link to any search tools that cannot be included in the common search interface—thereby providing convenient access to the range of resources listed above. One begins to glimpse here a world where information silos are a thing of the past, where library-provided content truly is presented in a unified, integrated fashion, where hypertext linking begins to realize its ultimate potential and users can customize offerings according to their preferences in a true portal environment. Luther provides a helpful overview of the technology's potential and some relevant links.²

These tools, whose function is also called “metasearching” or “broadcast searching,” are still in a relatively early stage of development. As with any other type of product they differ in capabilities, overall design, and cost. To fulfill their shared responsibility for selecting products that are affordable while offering the best value for the library and its users, reference personnel need to know as much as possible about the options. Common questions to consider would be: What is the purpose in acquiring the product? Is the tool needed? What does the user interface look like and how customizable is it? What type of end-user will it best serve? How does the tool actually work at a technical level? What will be the impact on bibliographic instruction? How would it impact reference staff in day-to-day work?

The pricing models offered by companies producing federated search tools have matured. Federated searching can be completely outsourced to a vendor or managed in-house. Outsourcing avoids the cost of purchasing and maintaining a server. Some vendors offer different tiers of service depending on the number of

database connections needed. It is important to carefully review library workflows, staffing resources and database connectivity needs before purchasing a tool. Although prices have dropped due to a more competitive business environment, implementation and maintenance can still be a substantial investment of money and staff time. Research to find the best tool for a given library is essential. One case the authors discovered in the course of contacting institutions about their experiences implementing federated search tools is instructive. A relatively small college purchased a federated tool after hearing several sales presentations. Once it was implemented, reference staff discovered that it conducted keyword searches, not the more sophisticated searches they were accustomed to doing on their OPAC. They also realized that because of the modest number of databases, there were few overlapping databases covering similar subject areas. For example, the typical patron searching business topics had no need to cross-search ERIC with the business databases. Ultimately the integrated search functions of the tool were disabled despite the investment in it.

The newness of these tools presents a major challenge to reference librarians to review and select. As noted, those currently on the market vary greatly in what they include and in overall concept and design. As Roy Tennant cautions in a *Library Journal* article, "The market is still in an early stage of development, with products varying quite a bit in features, stability, and ease of implementation. This means if you're purchasing a system, you should use due diligence in comparing features, talking with current customers, and making sure

that the product will deliver on its promises. It also means that implementation may not be a breeze.”³

Reference staff must be involved in the decision-making process when the library is considering acquiring a federated search system. It is, after all, the reference department as the front-line users and instructors who will be impacted the most once the system is deployed. Hence, the decision should not solely rest with the systems department or library administration. Reference staff needs to understand and contribute to the entire process from selection to implementation and beyond. They can help articulate the reasons why such a tool should be acquired, review how it operates, plan for impacts on workflow including reference desk service and library instruction, and of course try to see it from the end-user's perspective. The MetaLib review below contains in its discussions of the ResourceStore and administration/configuration issues some good examples of what reference librarians can and should contribute. It is our purpose here to alert reference librarians to these issues while highlighting some primary considerations in evaluating potential acquisitions.

This article will review some of these federated search packages. The reader should bear in mind that the authors were not able to test each product in an identical manner. Some have been on the market longer than others. Fully functional versions of some were available for review, while available versions of others were more limited. The quantity of published information on each varies considerably. Therefore, the differing length and level of detail of each review section is not to be taken as an implicit assessment of relative merit.

The following resources will be reviewed:

- MetaLib by Ex Libris
- ENCompass by Endeavor
- AGent by Auto-Graphics
- WebFeat by WebFeat & Thomson ISI

METALIB/SFX

NOTE: The authors were able to visit Colorado State University (CSU), a nearby institution that has implemented MetaLib. Our discussion incorporates some examples of how CSU has chosen to implement and customize various features.

The MetaLib system is marketed by Ex Libris (<http://www.exlibris.co.il/>), a company that previously was chiefly known for its integrated library system, ALEPH. MetaLib has a significant and growing customer base. Bundled with MetaLib is SFX, an OpenURL linking component. (OpenURL is a developing standard for dynamic, context-sensitive linking of electronic resources; it eliminates problems inherent to static linking by generating on-the-fly links to library-configured resources that are appropriate to the particular institution and user.) Those who implement MetaLib must also purchase SFX.

MetaLib is designed to serve as a portal to all library-provided resources by providing a single entry point to electronic and print collections—abstracting/indexing and other types of reference databases, full-text sources, library catalogs, digital repositories, selected fee or free web sources, and document delivery services. It is intended to work independently of the institution's ILS/OPAC so can be used regardless of ILS vendor. This gateway to local and remote resources overlays the

ILS and all other separate components of the library's access tools, allowing the user to remain within one common, familiar system regardless of what is accessed. Customizability that allows the user to personalize the system to best fit his/her needs is also an important part of the portal environment and is supported by MetaLib.

Basic Design and Searching Capabilities

There are two basic levels of searching: the Universal Gateway and the Information Gateway. The former supports federated searching of multiple sources such as reference databases and catalogs, along with displaying search results, using the MetaLib common interface.

To be included in this parallel search environment, a given resource must be configured. Successful configuration is a meticulous process and is dependent on external factors such as vendor support for communication protocols and the structure of database records. Most library catalogs currently are Z39.50 compliant and therefore can be configured for MetaLib searching, but not all database vendors or other information providers are. For those, other suitable translation protocols must be employed. The resources that are configured for the Universal Gateway offer a relatively simple search input screen, with two text boxes offering Boolean options. Available fields besides the broad "all fields" option are: subject, title, author, ISSN, ISBN, and year. Search input is translated into the appropriate syntax of each selected search target, sent to the target systems, and processed; the results are then returned and can be viewed within either the MetaLib interface or the native mode of the originating system. At

CSU, up to eight databases can be searched simultaneously. The library determines the number that can be searched together, but Ex Libris recommends eight as a maximum in order to keep loads on the system manageable. One can switch to the native search interface of any of the selected databases at any time.

Since MetaLib is intended as the common presentation of all library resources, the Information Gateway provides “link-to” capability for all sources to which the library wishes to provide access but which are not configured for the Universal Gateway. In this way the user is able to link to any desired source and is placed in that source’s native interface for search and retrieval. The link can lead to the home page or to a location within the resource, depending on what the resource itself allows. The Information Gateway thus serves as the resource discovery component, showing the patron what is available and providing linking to all included resources in the same way that the typical library without a federated tool does, but within the unified MetaLib menu system.

Tying together these two access layers is a resource catalog MetaLib calls the ResourceStore. This database comes prepackaged from MetaLib with a large set of existing resources. Each MetaLib installation is responsible for customizing this catalog to reflect the resources provided at that site. Also included in the ResourceStore is an array of broad subject categories intended to facilitate locating databases by subject. The subject arrangement can be customized to fit each library’s requirements. Complementing the structured subject access is a resource locator allowing keyword searching or alphabetical browsing of all available

databases. The keyword search can include terms from the database names/titles, descriptions, notes, and additional terms added by librarians to facilitate access and discovery. A given resource can be assigned as many categories or keywords as needed. Lists of databases all include an information icon (“i”) leading to the associated resource description; this prevents cluttering the screen while providing one-click access to the information.

Many of the screens and their content are customizable by the library. The subject access system obviously has to accommodate a great variety of libraries and collections, so it is designed to be tailored to each institution’s range of available resources, collection strengths, and preferred terminology. The screen appearance, including icons, can be customized to some degree. CSU, for example, created a simpler screen presentation. The subject search is presented as a “Pick A Category” drop-down box that displays the library-assigned subjects. Just below, two search boxes enable keyword searching of all databases by name (title only) and keywords (all words in the local ResourceStore record). Last is an “A-Z List” button for browsing all databases. Lists of available databases are displayed at center screen, with an icon for jumping to the native interface; those searchable via MetaLib also have checkboxes for inclusion in a federated search. Selection tabs allow display of “all resources,” “search resources” (Universal Gateway), or “link-to resources.” Quite a lot of basic information is included without the screen looking unduly cluttered.

Managing Search Results

Since searches via the Information Gateway are conducted in the native interface with the same look, feel and functionality as if one had entered the resource from outside MetaLib, results are presented with the original formatting, sorting, etc. as well. The primary value that MetaLib adds here is the availability of SFX OpenURL linking services. The Universal Gateway, on the other hand, attempts to integrate results displays as well as searching, although there are limitations to this capability as there are to the search function, given varying formats and record structures. In addition to linking to the sources pointed to by the database records, one can also refine an initial search based on the ensuing result set—either within that result set in order to narrow the retrieval by using cues in the metadata such as subject terms, or by using metadata within records to launch another search to retrieve more records beyond the initial set.

The results from a MetaLib search are subjected to a variety of conversion programs to convert the records from the native format to the unified MetaLib format, for the sake of consistency. Again, one can go straight to the originating database to view record displays in native mode, even before the system has finished compiling the results for all databases included in the search. Once the search is completed, clicking on the Merge button will combine the summary results into one set with the ability to deduplicate the record list. At CSU the maximum size of this set is 150; this limit is customizable by the institution. According to Ex Libris, the user can sort the merged list by author, title, or year, although that option

was not obvious to the authors during their test-drive. Several deduping algorithms are made available, with the institution being able to select the one deemed to best meet its needs. Each full-record display shows all databases from which the record came, again with the option to link to the original record. A merged set from a search on “welfare reform” in 6 databases and limited to 2003 returned 52 records that were deduped down to 32.

User Customization

A popular feature of the typical portal is the ability to personalize the resource selection and display in whatever way the user finds most helpful. MetaLib offers a variety of such options. Probably the key one is creation of one’s own list of the most-used resources, both databases and individual e-journals. It becomes part of the user’s profile and is the default list for searches. Saving queries for re-use, setting display parameters and formats for displaying and saving records, and establishing alerts to run automatically against selected resources at intervals chosen by the user are other handy options. Of particular note is the e-shelf, to which search or automatic alert results can be saved indefinitely. It thus functions as a personal catalog or inventory of useful information sources, to which users can add comments. The links in the records are preserved for later use.

Administration, Configuration and Maintenance

Any library choosing to acquire MetaLib and SFX can expect to spend a great deal of time on initial

configuration and setup. Configuration tables must be set up for each resource that is to be made searchable within MetaLib, and each SFX target must also be configured individually since the whole SFX concept entails targeting only those resources that are available to a given library's patrons or to defined categories thereof. The word used more than once in the literature to characterize the configuration environment is "meticulous." It can take months to do all the setup and configuration work necessary to roll out the system to the public.

Regarding impacts on staff, of the three MetaLib institutions whose configuration and maintenance staffing patterns are known to the authors, all use a team approach—not surprising given the amount of work and varying knowledge required. Typically, they include representatives from library systems/IT, technical services (acquisitions, cataloging/metadata), and public services. This is a good example of the breaking down of traditional boundaries of library work. In the metasearch environment, all functional areas are involved and must work together to ensure the tool meets the needs of that particular library's users and staff.

ENCOMPASS

ENCompass, produced by Endeavor Information Systems (<http://encompass.endinfosys.com>), vendor of the Voyager integrated library system, is an integrated series of XML-based tools designed to aid libraries in becoming "digital libraries." (XML, Extensible Markup Language, is a meta-language that allows one to build descriptive documentation. XML documents are a series

of text and descriptive tags. The tags, called markup, place a tree structure on the document and allow one to describe the content of the document to be displayed. XML is designed to describe and communicate electronic information.) It combines a federated search function (“ENCompass for Resource Access”), a digital asset management system (“ENCompass for Digital Collections”) and an OpenURL resolver (“LinkFinder Plus”). All three are sold separately or as a complete package.

The federated search component includes a comprehensive range of resource types: Abstracting/indexing databases, e-journals, e-books, websites, and OPACs. Endeavor has recently partnered with the FAST search engine to enhance response time.

To test the federated search side of ENCompass, the authors used the University of Kansas’s (KU) implementation. Accordingly, the authors viewed the product in terms of the institution’s value-added customization; it may look quite different at another site. It was not feasible to arrange full access at a variety of sites, but we were able to gain some insights into basic functionality.

Basic Design and Searching Capabilities

The federated search interface and screen layout are almost completely customizable by the library, so we will not address the screen design or compare it to other systems. Suffice it to say that a primary consideration in choosing a federated system is likely to be the ease and extent of customizing the presentation into one that satisfies the library and, by extension, its users. KU’s

main search screen follows the apparent practice of most libraries with federated tools in presenting databases grouped by library-defined subject areas (Arts and Humanities, Business, Current Events and News, General and Reference, Science and Engineering, Social Sciences, etc.) plus a “Popular Databases” category accessing five general, multidisciplinary databases. An “all Digital Library Databases A-Z” category (“Digital Library” is KU’s name for the ENCompass service) enables searching all federated databases at once. Each category has a checkbox which, if selected, causes the search to be run against each database in the category. Alternatively, one can expand a category by clicking a button beside the category checkbox; this displays a list of all databases within that category, each with its own selection box. Multiple categories can be expanded simultaneously. The user must check a separate box in order to include the KU libraries’ catalog in a search. Links to descriptions of each individual resource, as well as each subject category, display next to the entries. In keeping with unified resource access, an A-Z list of all the library’s electronic resources is at the bottom of the screen with a note indicating not all electronic resources are available in the Digital Library.

There is no technical limit to the number of resources that can be searched together—and KU has not set one—although in a practical sense the number selected can impact response time. If a number of remote databases are being federated, response time can vary widely depending upon the speed of the university’s network and the speed/connectivity to the remote server. The number of results to review is an obvious factor. For a highly specific, focused search, however, this is a

desirable capability. Even when a broadcast search of all 40 databases at the KU site was conducted, response time was deemed acceptable. While a search is in progress, the screen refreshes every few seconds and a red flashing “searching...” indicator is prominently presented, reassuring the user that the system is still responding.

The default search is a simple input box beside a drop-down menu box for selecting the available search fields: free text, global keyword, title keyword, author keyword, and subject keyword. No Boolean combinations are possible in simple search. The advanced search screen provides three boxes with full Boolean capabilities between the boxes; additionally, each box has a drop-down menu beside it with options for “any of these” (OR) which is the default, “all of these” (AND), and “as a phrase,” allowing definition of logical relationships between multiple words within each box. The same search fields are available with the exception of “free text.” In many resources a phrase search is actually not possible, in which case it reverts to AND logic. The lack of capability for more sophisticated searching is to be expected; it is endemic to searching across disparate sources regardless of the quality of translation protocols and field mapping. Search history, including options to re-run or edit previous searches, is offered. No limiters are available. A quick navigation box lets the user easily jump between search options, existing results displays, and other available actions.

Managing Search Results

The format of search results is customizable. As delivered to the customer, ENCompass incorporates a standard display similar to MetaLib. KU chose to customize displays in ways that preserve the appearance of the native source. After linking to a full record from the initial brief-record display, one can page forward or back through the full records. In both full and brief records is a prominent button linking to the native interface for searching there. KU has implemented Endeavor's LinkFinder Plus as KULink. The link resolver sends the user to a links resolution screen where one can connect to the full-text display. If the article is full-text one can avoid going through the intermediary screen and launch directly to the article via the evoke command.

After a search completes, a "results summary" screen appears that lists the number of records for each database. One can view records grouped by database, or click a "display all results" button to see a merged set. There is no apparent limit to the number of merged records that can be displayed, although a maximum of 50 is retrieved from each individual resource in most cases. It is not clear what criteria are used to determine which 50 items get displayed—relevance ranking is not used. Results sort alphabetically by title, including initial articles. However, this is not all bad since deduping is not possible at this point and the title sort makes duplicates easy to spot. User-defined sorting by date as well as the default title order is available. Records cannot be marked for later use, but this may become

possible with future upgrades or when user personalization features are activated.

User Customization

Endeavor offers user login and associated features such as creating a personal resource collection for the search screen, saving searches, and saving individual records for printing, emailing, etc. (see preceding section). The user customization has not been implemented at KU, and it is not clear how many sites have already done so.

Administration, Configuration, and Maintenance

Any library choosing to implement ENCompass for Resource Access or Digital Collections and LinkFinder Plus will be spending a certain amount of time configuring the system. It should be noted that ENCompass as a suite is a vast tool. Both Kansas and Cornell University have teams devoted towards implementing their projects. Cornell was an ENCompass development partner and they state that presumably they received different levels of training than a “typical” customer. The team at Cornell had to work through many issues on their own and describe themselves as “self-taught.” They found the learning curve to be “fairly steep and long.”⁴

AGENT

NOTE: The authors tested a demo database provided by Auto-graphics, using the Toronto Public Library’s

implementation of AGenT entitled “All in One Search” and also JerseyClicks, a statewide gateway.

AGent by Auto-Graphics, Inc. (http://www4.auto-graphics.com/product_agent.html) is best known as a provider of large, federated, statewide library union catalogs. The customer base includes the states of Wisconsin, New Jersey and the Texas Education Agency’s Texas Library Connection. Auto-Graphics does not offer an OpenURL link resolver at this time, but rather works with a number of existing resolvers already on the market.

AGent can be used by a single library or by a consortium to access and make available a locally installed ILS as well as the catalogs of other libraries from one interface. A&I databases as well as selected websites can be searched using this interface. Also searchable are proprietary in-house databases and repositories of Word and PDF documents. Auto-Graphics’ marketing literature suggests that a library can start out federating other library databases using the Z39.50 protocol and expanding functionalities as budget allows.

Basic Design and Searching Capabilities

AGent’s search screen allows databases to be grouped. For example, the catalogs of a library consortium could be a group, and databases on a specific subject such as business could be another group.

The Keyword or Simple Search searches all indexes (Title, Author, Subject and Notes). Keywords can be entered in any order; multiple words are ANDed. The

Advanced Search is used to create “search queries” with full Boolean capability. This can be done with up to three keyword combinations, across multiple search headings. Within each keyword combination is the option to search all headings. End-users also have the opportunity to narrow the search by choosing from other drop-down box options such as searching specific fields. The system also allows searching a variety of standard publication identification numbers including: ISBN, ISSN, LCCN, and a variety of GPO numbers including: SuDoc, Item, Order, Report, Monthly Catalog, and Ship List. The Original Control Number enables the user to search the number located in the MARC 035 field or Publisher Number which are numbers usually assigned to sound recordings, video recordings, printed music or other music-related materials.

Managing Search Results

Search qualifier options allow the researcher to narrow the search’s scope. Search Qualifier is used with the Advanced and Research search options.

Result “sort by” qualifiers can be preset in Advanced Search. These include sorting by newest title first, newest title last, alphabetically by title, and alphabetically by author. Other narrowing qualifiers include year of publication, language, and media. Form qualifiers include limiting the result set to Braille, electronic, large print or microform. Another qualification option is to search by miscellaneous. This option appeared to allow limiting by item type such as full-text URL, government publications, and thesis. There did not appear to be any ways to save and re-execute a search.

A results summary screen lists the number of records for each database. Databases return results in no particular order. Results can be grouped by discipline such as All Basic Research, All History, All for Students and Homework Help, All Business, etc. The results display can be configured to meet the needs of a specific library audience. In some installations feedback in the form of a traffic light indicates the success or failure of the search against a database. Next to each database is a traffic light showing red, yellow or green. When the databases successfully return the completed set from each specific database the light turns green. If there was an error during the build the light is red and an error message is displayed. Results sets can be viewed as soon as the returned database names appear in the results list. The return of result sets in AGent was very fast. Both the simple and advanced search allow the user to further modify the search, start a new search or start over within the current set of selected databases. Records are grouped by database. There is no option to merge and view result sets. The results display in their native "look and feel." There did not seem to be a limit to the number of search returns results.

User Customization

AGent has a number of personalization features. My Account allows the user to change password or PIN number. Users can change or edit their personal data, such as e-mail address. Within My Account patrons can view My Search History, which stores the last twenty searches in reverse chronological order. The search history shows the type of search performed (keyword,

browse or advanced) as well as the search terms entered. Searches can be modified, stored and re-executed on a regular basis.

My Favorite Search source is a MyAccount feature that remembers the databases that the researcher used on previous searches. Favorite frequently used databases are indicated as being "on" by being pre-checked in the list of choices. Changes to the favorites do not take place until the user logs in again for another search. At this time there is not a way to mark, save, print or manipulate result sets.

Lastly, My Preferences allows researchers to customize a number of system options according to their need. A default display language can be selected from the languages menu. Other customization features include defining the way search results are displayed. A few of the many options available include: 1. The number of records returned can be set to 10, 20, 30, 40, 50, or 100; 2. Sorting results by the criteria listed in the preceding section; 3. Default format for full record display (Card, Labeled or MARC); 4. Number of searches to save in search history--this can range from 0-50; 5. Default search method including Keyword Search, Browse or Advanced Search.

Administration, Configuration, and Maintenance

AGent has an easy-to-use web-based systems administration (sysadmin) module. Colors, banners and most of the "look and feel" changes can be managed from this module. The end-user search screen is set up using this tool. The system includes group name

management allowing the aggregators or resources to be ranked so they show in the order on the screen.^{5, 6}

WEBFEAT

NOTE: The University of Illinois at Chicago was an early WebFeat customer and the first library to implement WebFeat 3, the current version as of early 2005. The authors arranged full access to test this version.

WebFeat (<http://www.webfeat.org>) originally was marketed primarily to large public libraries systems, most of which have multitudes of disparate databases, but was also targeted to corporations. WebFeat partnered with Thomson ISI to provide a federated search tool for their expanding product portfolio. This gave WebFeat an entry into the large academic market. While it is a stand-alone federated search tool, WebFeat is compatible with the major OpenURL link resolvers, so no bundling is involved in a purchase.

WebFeat's Todd Miller believes that federated search engines should be offered to customers as an ongoing service rather than a software product. Unlike most of his competitors, WebFeat offers a try before you buy program. This is a refreshing change from vendors who send in a representative on a sales call who conducts a preplanned canned demo that works flawlessly, and then expects the library to purchase the product without actually being able to test it in their local environment. Libraries purchase a subscription to WebFeat's services, which include building the custom interface used by the public. The company also develops and maintains the numerous custom "translators" that enable federating the various databases. A subscribing library simply develops

a list of the resources that it wishes to federate and WebFeat builds the interface, maintains the translators and develops the end-user interface.

Basic Design and Searching Capabilities

The WebFeat interface at the University of Illinois at Chicago (UIC) is called qUICsearch. UIC set up the “look and feel” of their interface to complement that of their library homepage. Not only does WebFeat support library branding in this way, it is also highly customizable in every aspect, with a great range of implementation options.

The initial screen displays a list of broad subject categories headed by the “All Resources” option invoking the complete alphabetical list of 117 federated databases, which persistently displays below the category list. Arts & Humanities, Health Sciences, Multidisciplinary Resources, Science & Technology, and Social & Behavioral Sciences are the other available categories, each having a checkbox for selecting the entire category. Each broad category can be expanded by clicking beside the box, which then displays specific disciplines—for example, Criminal Justice, Education, or Sociology under Social & Behavioral Sciences--that can in turn be selected by clicking one or more boxes. The selections automatically check the boxes beside each applicable resource in the A-Z list. The user can also select any desired resource using that list without having to choose a category first.

UIC’s main search screen provides three boxes with all Boolean options. Following typical practice, entering multiple terms within a box defaults to AND. Available

search fields are labeled as All, Keyword, Title, Author, Abstract or Subject. The default is set to Keyword. Although we could not obtain absolute confirmation, we suspect that Keyword searches Subject, Title, and Abstract fields simultaneously whereas All is indeed all-inclusive. Available limiters are publication date (all available dates or a range of years may be specified) and full-text articles only. WebFeat apparently supports limiting to peer-reviewed articles as well, but UIC does not offer it. Considering the tiny proportion of available databases that actually offer this limit, it would seem to be of little use in any case since peer-reviewed articles from all databases not featuring this limit would be eliminated.

New to the current version is a thesaurus function containing both LC and MeSH headings. Currently the thesaurus cannot be used in an initial search; it is invoked by a link titled "Find related terms" at the top of any search results screen. A simple alphabetical list with the closest match to input terms appears in a pop-up box; only one heading can be selected. LC and MeSH headings cannot be separated. UIC is hoping to be able eventually to separate them, link MeSH only to the Health Sciences category, allow starting a search using the thesaurus, and enable choosing more than one heading. While not well developed at this point, this certainly is a potentially significant enhancement.

WebFeat presents a simple, clean search interface with a less busy appearance than MetaLib or ENCompass. As one would expect, however, there is a price to be paid for simplicity. All the pitfalls presented by using the same search strategy across a disparate set of databases come into play, a problem that is not unique to

WebFeat. With the implicit Boolean AND, especially with full-text searches being executed in many databases, one encounters the usual problem of poor precision when the search terms are far apart. Restricting a search to the subject field does not necessarily mitigate this problem; the field-mapping algorithm appears to cause a subject search to default to the entire record, including full-text, when no subject field exists in a given resource. On the plus side, response time is very good. We did several searches against all 117 resources, and none took more than two minutes to complete.

Managing Search Results

The default order of results is by database of origin (“grouped” in WebFeat parlance). They sort in roughly alphabetical order, but the speed with which results are returned also influences the display order. Within each database display, items are sorted in reverse chronological order. All databases selected for the search are listed at the top of result screens, including those with zero hits. Number of hits is shown for each, and the resource description displays on the right. Each resource name is linked to enable jumping to the results for that resource. The sort order can be changed to: 1. Article title (includes initial articles such as “the” in the sort order); 2. Author (of limited use since it sorts by whatever letter occurs first, which can be either first name, last name, or “Anonymous” as well as “by” which sorts with B); 3. Date; 4. De-duped which sorts by normalized title and seems quite reliable; 5. Selections only, if items have been checked; 6. Relevancy ranked (according to search terms’ occurrences in titles); and 7.

Merged, which sorts in the same order as the “grouped” display but is numbered continuously instead of starting over with each database group.

The default format is brief records. The library sets the maximum number of records displayed per database. When the number of hits exceeds this limit, the next set of records is linked at the end of the list. A link titled “View” leads to the full record, which always opens in a separate window and displays the record in native format. An OpenURL link button also is available. The native interface in the new window also accesses the search engine for that database where additional searches can be executed or linked full-text viewed. Returning to the brief results list is as easy as closing the window. Results can be selected for printing, emailing, or saving to a file.

User Customization

The latest version introduces a “My WebFeat” feature supporting personalization by the user. It enables building custom database menus, subject categories, and automatic alerts to capture newly added records matching saved searches, as well as saving and re-executing saved searches whenever desired. Academic sites are expected to begin implementing My WebFeat in mid-2005. Exporting records to bibliographic management packages such as ProCite, EndNote, and Refworks is also possible.

Administration, Configuration, and Maintenance

WebFeat is a completely outsourced service. The company handles all configuration, maintenance, and upkeep of the “translators.” The “look and feel” of the search screen interface is customized by WebFeat for the requesting institution based on their input and specifications.

One problem with federated search tools is that library staff must work very hard to set up and maintain the interface’s “look and feel,” and establish/maintain the connection infrastructure to the external or internal databases that enables them to be searched. In some cases multiuse, bundled tools that are marketed as an integrated interface also depend upon outsourced companies for the maintenance of their connections. At times these outsourced relationships are complex and can result in non-connectivity issues. Development and maintenance of the translators is one of the WebFeat’s strengths. WebFeat’s staff has developed over 1200 pre-built aggregated database “translators.” For a fee, additional custom “translators” to accommodate localized databases can be built upon request.

In Todd Miller’s article “WebFeat: The Boutique Aggregator,” he states, “It (WebFeat) assumes a world where the roads have been built and paved but they lead to many different destinations and local customs. Additionally, WebFeat assumes that, in the foreseeable future, it is unlikely that standards are going to be successfully imposed on the locals.”⁷

CONCLUSION

Federated search tools are proving to be a useful addition to the reference librarian's toolbox. Since they rely on a limited set of common attributes across disparate databases they also have severe limitations which preclude viewing them as a catchall solution for precision searching. In their current state of development the authors see them as a starting point for those end-users who are not sure where to begin in seeking research sources. After reviewing their search results they can always go directly to those databases that appear to have the most material of interest and drill down into more detail. Federated searches seem especially good at facilitating this ability.

A fundamental principle to keep in mind is that the more databases a tool can cross-search, the less precise the results are. This is a major downside to all these systems. They are only as precise as the translators they employ. All databases being federated must have some common searching fields; these are defined, translated as needed, and built during implementation. The fewer fields shared in common by all databases, the poorer the search precision. Most federated searches are keyword searches, most often in title, author, or subject fields. Phrase searching is often problematic or simply unavailable. Limit features, such as restricting article searches to scholarly or full-text journals, are seldom fully functional for these reasons. Reference librarians will need to emphasize the limitations as well as benefits to their users, and use judgment as to when users would benefit from a federated search and when

they should use a stand-alone database. For an inexperienced user familiar with the large result sets of general search engines, the typically large and imprecise retrievals characteristic of federated searching may look quite focused by comparison. This perception can perhaps be used as a starting point to guide users toward more sophisticated search instruments and a better understanding of how a library can benefit them—what Zimmerman calls the “teaching moment.”⁸

At present, federated searching is likely to work well for many undergraduates and members of the general public who need some information. For those at a research or specialist level, it does not begin to emulate the sophisticated capabilities (such as thesauri and limiters) of the native interfaces of specialized databases. Many of these users will, for the time being at least, continue to be best served by “going native” unless they are seeking background information outside their areas of specialization.

It is beyond the scope of this paper to make detailed evaluative comparisons between the systems under review. Each combines functions and resources in its own unique ways. Customizability, one of the strengths of federated tools, presents a challenge to review and compare. Instead of assessing one interface, libraries will likely need to compare several implementations of each system. Focusing on existing installations in institutions similar to those of the potential buyer should help to mitigate this challenge.

In all cases, each library will need to make careful judgments based on solid research. Exactly what one wishes to accomplish with whatever system is adopted is,

as always, the driving factor. Representative questions that may be considered include:

1. Who is the target audience of the library and ultimately the software? (The general public, undergraduates, graduates, faculty...)
2. Will it be used only to federate multiple databases, or are additional functions--such as context-sensitive linking via OpenURL, implementing a digital asset management system, or integrating institutional repositories into the package--also desired?
3. Will the majority of the target databases be external (i.e. subscribed databases such as those offered by OCLC, EBSCOhost, Gale Group, and ProQuest) or internal, created by the library or its parent institution?
4. Is there a plan to use the federated tools in a consortium? (On campus or multiple academic or public sites?)
5. How is the tool bundled? If a federated search tool is purchased does the library also have to purchase an OpenURL resolver as part of the package?
6. How will the tool affect library workflow? Is there enough staff to implement and maintain a federated search tool or should the library consider an outsourced tool?
7. How much configuration does the library want to do? What impacts will such a tool have on the reference or systems staffs?
8. What can the library afford?

More generally, how effective are these tools at accomplishing their stated purpose of simplifying information discovery, search, and retrieval? The intent is to reduce or eliminate barriers to effective library use.

At the same time, the hope is that library-provided resources in all their diversity will become more visible to the user and hence better utilized—creating more synergy among the resources and helping libraries to justify the considerable investment they make in physical and virtual collections. But what is the potential for realizing these goals? While some tools are more “mature” than others, all have been on the market for such a brief period that it is too early to draw definitive conclusions.

This review has, however, led to some general impressions of their utility at this time. The resource discovery component is the most advanced in development. The tools reviewed here all do fairly well at bringing together in a unified presentation the institutions’ available electronic and print resources. This goes far toward eliminating confusion over where to find desired information by erasing traditional distinctions between discrete categories of library resources. Although a common search-retrieval-display interface is unavailable to those users needing resources that cannot be configured for federated searching, all resources still are more easily located. Different systems and varying local implementations accomplish this to differing degrees depending on interface design, but in principle the concept of a gateway or portal is being realized.

The true unified search and display of disparate resources is much more challenging and hence not fully developed by any means. As we have seen, any tool is limited by the configurability of the target resources, which in turn is highly dependent on their design/structure and the technically complex translation protocols currently needed to achieve an integrated

presentation. On the other hand, given that multiple vendors are competing for customers and that data interchange standards continue to evolve rapidly, every indication is that the inherent limitations to federated searching will be gradually ameliorated over time. While comprehensive yet simple one-stop-shopping is probably a long way off, incremental improvements can be expected on a continuous basis. These may accelerate if the Metasearch Initiative is successful. Launched in 2003 and sponsored by the National Information Standards Organization (NISO), it is a working group of library and industry professionals attempting to “identify, develop, and frame the standards and other common understandings that are needed to enable an efficient and robust information environment.”⁹

COMPARISON TABLE

	MetaLib	ENCompass	AGent	WebFeat
Digital Asset Management System		X		
Open-URL	X Ex Libris offers SFX	X Endeavor offers LinkFinder Plus	Works with multiple vendors	Works with multiple vendors
Bundling Requirements	X			
Number of Databases that can be Searched Simultaneously	No limit* 8	No limit	No limit 5–10	No limit
Deduplication	X			X
Native Search	X	X	X	X

Interface Available				
End-User Customization	X	X	X	X
Advanced Search		X	X	X
Boolean Search	X	X	X	X
Phrase Search**	X	X	X	
E-Mail Results	X	X		
Search History	X	X	X	

* Vendor recommended limits

** Does not work in all databases

REFERENCES

- ¹Roy Tennant, "Library Catalogs: The Wrong Solution," *Library Journal*, February 15, 2003, 28.
- ²Judy Luther, "Trumping Google? Metasearching's Promise," *Library Journal*, October 1, 2003, 36-39.
- ³Roy Tennant, "The Right Solution: Federated Search Tools," *Library Journal*, June 15, 2003, 28-30.
- ⁴Cornell University Library, "FAQ: About Cornell's ENCompass Development Project," http://www.library.cornell.edu/cts/encompass/ENC_FAQ.htm, (accessed July 29, 2003).
- ⁵Agent Powers Jersey Clicks, *Library Journal*, June 15, 2004, <http://www.libraryjournal.com/article/CA423770.html>, (accessed May 20, 2005).
- ⁶JerseyClicks: New Jersey's One-Stop Library Gateway to Quality Information, <http://www.jerseyclicks.org>, (accessed May 21, 2005).
- ⁷Todd Miller, "WebFeat: The Boutique Aggregator," in *National Online 2001, 22nd: Proceedings* (New York: Information Today, 2001), 249-256.
- ⁸Devin Zimmerman, "Metasearching's Teaching Moments," *Library Journal*, September 1, 2004, 54.
- ⁹Andrew K. Pace, "Technically Speaking: Much Ado About Metasearch," *American Libraries*, June/July 2004,