Scherlen, Allan (Balance Point column editor) John Boyd,, Penny Pugh, Marian Hampton, Pat Morrison, Frank Cervone, **"The One-Box Challenge: Providing a Federated Search that Benefits the Research Process"** ***Serials Review***. 32:4 (Dec. 2006) 247-254. Copy of record available from Elsevier ScienceDirect: doi:10.1016/j.serrev.2006.08.005

The Balance Point

The One-Box Challenge: Providing a Federated Search That Benefits the Research Process

Allan Scherlen (column editor), Collection Development Librarian for the Social Sciences, Appalachian State University, Boone, NC 28608, USA

Contributors:
John Boyd, Reference and Instruction Librarian, Appalachian State University, Boone, NC 28608, USA

Marian Hampton, Coordinator of Library Instruction, Hillman Library, University of Pittsburgh, Pittsburgh, PA 15260, USA

Patricia Morrison, Off-Campus Resources Librarian, Grossmont College, El Cajon, CA 92020, USA

Penny Pugh, Head Reference Dept., Downtown Campus Library, West Virginia University, Morgantown, WV 26506, USA

Frank Cervone, Assistant University Librarian for Information Technology, Northwestern University Library, Evanston, IL 60208, USA

Available online 28 September 2006.

Abstract
The one-box federated search of databases brings as many challenges as promises to database searching, especially in terms of adapting these systems to user needs and the effects this new mode of searching will have on users' research behaviors. This issue of "The Balance Point" presents several librarians with strong interests in reference and instruction, who tell the story of adapting a federated search system for their libraries and reflect upon how federated searching can change the way students do research and on the implications federated searching has on information literacy skills and the quality of results found.

Editor's Introduction

“Just give me a single box that searches everything,” a friend of mine in the political science department said to me not long ago while we drank coffee and discussed his research. He was overwhelmed with what he saw as too many library databases that seemed impossible to keep up with. “Hope is on the way,” I assured him. “We are looking into getting a federated search engine that is supposed to do exactly what you are asking.” I may have been going out on a limb with that reassurance since our institution has yet to settle on a federated search product that works the way we want. But one can easily sympathize with the frustration of researchers confronted with numerous databases with various interfaces and features.

No doubt, my friend's frustrations are shared across academia because federated search products are emerging and developing at an impressive pace. In turn, librarians face a challenge: deciding which federated search product works best for their institution, how to adapt the system and what effects federated searching will have on their users' research process. Federated searching will certainly make some aspects of research easier, but will it make it better?

For this issue of “The Balance Point,” librarians tell the story of their search for the perfect federated search product for their library and reflect on how this new mode of searching is affecting how their users do research. John Boyd, reference and instruction librarian at Appalachian State University, relates how his library looked extensively at CSA's still somewhat beta product, MultiSearch and finds that it presents as many problems as benefits to student users. Marian Hampton, coordinator of library instruction from the University of Pittsburgh, describes the adoption of a WebFeat product that they branded as “Zoom.” She makes poignant observations about the “good enough” results of federated searching and the continued need for encouraging information literacy skills. Patricia Morrison, an off-campus resources librarian of Grossmont College Library, takes us through the ups and downs of their version of Serials Solutions' Central Search, noting some overlooked bonuses of federated searching, such as its usefulness to interlibrary loan staff. Penny Pugh, head of the reference department in the downtown campus library at West Virginia University, tells the story of implementing the WebFeat product they named E-ZSearch. Finally, Frank Cervone, the assistant university librarian for information technology at Northwestern University Library, describes the complex issues in any federated search implementation process and how usability testing is critical to getting the system to work well for users.

Formal evaluations of federated search products can readily be found online and in the library literature. This issue of “The Balance Point” provides a different perspective to the evaluation literature: a balance of narratives by mostly public services librarians who recount the process of acquiring a federated search product as well as provide some thoughts on the implications federated searching will have on their users' research behavior and the quality of that research. These perspectives dovetail nicely with the past issue of Serials Review which had primarily a vendor focus on federated searching (v.33, no. 3). Their shared insights bring welcomed flavor to the evaluation literature.

Appalachian State University Library's Quest for a Federated Search Product
John Boyd

Two years ago librarians at Appalachian State University first examined federated searching and concluded that the design and functionality of the products did not warrant the high cost. At that time we looked at Innovative Interface's MetaFind and the WebFeat Prism interface. The primary purpose of finding the ideal federated search product was to give our users a simple interface to multiple resources. We wanted them to discover the neglected resources that libraries spend great amounts of money on and that are valuable to their research.

This past year we embarked on a trial of a relatively new player to federated searching, CSA's MultiSearch which was launched in the summer of 2005. During that trial we also examined two other federated search products, WebFeat Express and Serial Solutions Central Search, with plans for trying more federated search products in the months ahead. So, based upon our experience with recent trials, I understand the value of carefully examining as many federated search products as possible before determining which product to purchase.

Since the CSA product is so new and since the others we have tried have been described at length elsewhere, the focus here will be on MultiSearch's interface and functionality. MultiSearch uses the CSA Illumina platform, and for those libraries subscribing to CSA databases and whose students are familiar with the Illumina interface, there may be a strong incentive to select this federated search tool to complement the full suite of CSA databases.

The default setting of the initial screen in MultiSearch is a Google-like single box in MultiSearch, called QuickSearch, which allows the user to search by keyword. Like Google, quotation marks are required for a phrase search. This present as an advantage for students expecting Google-like functionality. Below the search box the user has the option to limit a search by subject areas (predetermined by the library in the Administration Module) from a pull-down menu and by date range. There is also an option to select “Specific Databases” to search. Above the single search box, along with the QuickSearch tab, are the Advance Search and Search Tools tabs. Even for those of us in our library who were previously unfamiliar with the Illumina interface, the design is clean and attractive.

The one-box QuickSearch looks for words in any field of a database record, dependent, of course, on the fields available in a particular database; the Advance Search provides the user the option of selecting specific fields to search. The Advance Search screen presents three rows of search boxes, with a default search of “anywhere.” The user can also choose other fields to search, such as author, title, subject, or keyword. Located under the Search Tools tab is a link to “My MultiSearch,” which allows the user to create a personal profile for saving searches, creating e-mail search alerts, and saving search results, and preferences on how search results are displayed and sorted, including fastest first and relevance rank.

Upon entering a search, a user will be presented the message “Status: in progress,” which will change to “Status: complete” once the search is done. Databases being searched will either be displayed (expanded list setting in the Administration Module) or not displayed (collapsed list in the Administration Module). The only indication that the search is still in progress is the in progress message, which can easily be missed. CSA tech support assured us they will provide a more visible indication, such as the commonly used hourglass or other graphic. If the default setting is set to the expanded list, the user will see the number “0” for results listed by each database until the final number of retrieved results appears. Impatient students might interpret the apparent result number “0” to be the final results rather than an indication that the search is still in progress and decide to move on.

A federated search engine that has a slow response time may not impress students, who are accustomed to the fast speed of searching the Web, the library catalog, and individual library databases. We can expect students to want a federated search engine that compiles a list of results quickly with the most relevant results retrieved first; however, slow searching may not be a problem unique to MultiSearch but inherent in federated search engines—the speed of the slowest database determining the overall speed of a federated search.

Then there is the problem of getting results that are most relevant first. Searching within the seventeen social sciences databases included in our CSA trial, the results of almost every search we performed yielded citations from JSTOR or Project Muse at the top of the list. While the prominence of these two databases in every set of search results may be a tribute to the quality in their response time, I suspect that other factors may be involved. For some reason, more relevant citations from other databases consistently appeared farther down the list.

This brings us to the most frustrating part of MultiSearch, the “Sort Results” option. MultiSearch provides the option of sorting which is important because results are initially displayed in a less than satisfying arrangement as they are retrieved from each database. In addition to “relevance” and “reverse chronological order” (most recent first), there are sorting options for “alphabetical order by title,” “author sorting by author's first name,” “author/title by the author's last name, then by title,” and finally, “alphabetical order by the database name.” These last four sorting options are curious, to say the least. But one may wonder even more whether a user will even use the “Sort Results” option at all. A user is more likely to just begin scrolling down the results list.

As a reference librarian working with students in the classroom and in the reference area, I sympathize with their frustration at having to search multiple databases, each with its own interface and idiosyncrasies. Many students are perfectly happy to search a general aggregator database and, with a few full-text articles in hand or in a flash drive, believe they have adequately researched their topic. Often a search of one general database will suffice. But if we can help our users easily broaden the scope of resources employed for

their research by introducing a user-friendly metasearch tool that provides relevant results quickly, so much the better.

The University of Pittsburgh's Federated Search System
Marian Hampton

The University of Pittsburgh's University Library System (ULS) has long prided itself on being at the forefront of providing a robust collection of databases, e-texts, and e-journals in response to library user demands for desktop access to library resources. By 2003, however, usage statistics and usability studies revealed that patrons were having difficulty selecting and navigating the hundreds of online resources that the ULS offered. The ascendancy of Google as the undergraduate resource of first choice, with its simple interface and quick return of results, basically set the expectations of what online searching should be. The ULS sought some way to provide a unified interface for searching our online collection, and the emerging metasearching or federated searching technology seemed to be the answer. But if the ULS was to enter into the costly and time-consuming endeavor of implementing a federated searching system, it needed to offer more to its patrons than just a simple unified search. The question was whether any of the new companies offering federated searching was willing to work with the ULS to develop a tool that met all the demands of both Web savvy undergraduates as well as disciplinary experts.

ULS public services and systems personnel closely examined several metasearching tools. ULS personnel ultimately chose WebFeat because the product could federate across every online resource that Pitt offered and the WebFeat business model offered a hosted service rather than software requiring backend configuration and maintenance. The WebFeat team was very open to working in partnership with ULS personnel to design a user-friendly interface and essentially offered the design team a blank canvas. The ULS requirements were that the interface had to be easy to use, had to effectively search across and confederate all online resources the ULS offered, and be able to work in conjunction with the current e-serials management service. It also had to be able to deliver this service at any distance to accommodate Pitt researchers around the world. Interface development was approached from a public service and end user perspective, employing data from past usability studies to inform the process.

The design team worked to construct two interfaces: a basic interface that users first see on the Pitt Digital Library home page and an advanced interface designed for more in-depth disciplinary searching. Both basic and advanced interfaces are available from an alphabetical list of databases as well as all disciplinary database collection pages, allowing users to select the best method for their needs. The basic interface on the Pitt Digital Library home page offers users the ability to quickly search across both the library's catalog and a pre-selected collection of the library's largest databases.1 The advanced interface allows experienced searchers the ability to select the databases they want and search them by keyword, title, author, abstract, or subject and limits results by date range.2 Result lists offer brief citations and links to the full records and all full-text retrieval options.

The new interface was branded “Zoom” to highlight the ease and speed of its searching, and the ULS committed significant marketing effort and funds to make sure patrons became aware of the new service. The ULS received the 2006 WebFeat President's Awards for Innovation in recognition of the development of the Zoom system and its innovative marketing of the new service. In the two years since introducing Zoom, the ULS has seen over 300,000 full-text retrievals via Zoom. Our unpublished usability studies have shown that participants frequently used Zoom as their primary search tool, with over 50 percent of respondents indicating that they knew and used Zoom when searching library resources.

Roy Tennant's statement that librarians like to “search” and everyone else likes to “find” rings true in this situation.3 Federated search systems are popular due to the rapid growth in the number of online resources and the huge variation in interfaces. Until recently, users had to learn the rules and peculiarities of each resource in order to become proficient searchers. The benefit of metasearching is obvious—one simple interface for several sources and immediate access to full-text retrieval. Librarians, however, are concerned that federated searching presents an illusion that information is simple to find. As expert searchers, we recognize the great breadth and complexity of information and have learned specialized skills to manage these issues.

Librarians see federated searching as providing “good enough” results, but not delivering what we feel users really need. Those unsatisfied, confused, or frustrated users who come to the reference desk for help after metasearching are just a fraction of those using federated searching, and these people may be just a fraction of a greater group who may not even understand what they are missing in terms of quality results. Yet, a growing number of patrons have used federated searching enough to become successful and skilled searchers with the tool, so we cannot simply fault the tool when students who use it are unsuccessful.

Federated search systems are in the early stages of development, and will, no doubt, improve and adapt to user demands over time. While these systems may not yet deliver what every librarian would hope for, they do get users closer to the resources they may have missed but should be relying on for their research—something we have been struggling with ever since Google arrived on the scene. The seeming efficiency of federated searching does not replace the need for such basic information literacy skills as crafting smart search strategies or evaluating results. The challenge for public service and instruction librarians is to encourage users to look beyond “good enough” results and take those next steps toward developing sound searching skills and critical analysis of results.

The Impact of Implementing Federated Searching at Grossmont College Library
Patricia Morrison

One thing I know for sure about Serials Solutions' Central Search is that it is much easier to troubleshoot interlibrary loans with a federated search engine!4 This perhaps requires some explanation. When our students at Grossmont College find a citation to an article in

one of our many databases, they can complete an online form and send it to our ILL technician to initiate an interlibrary loan; however, if that form is missing vital information, our interlibrary loan technician hands over the incomplete form for me to complete. Trying to guess which of our many databases it came from was difficult. Often, I would “Google” it because the citation might appear in someone's bibliography on the Internet. Now, however, completing that citation is a snap! Since the requestor usually gets the title of the article right, we can use Central Search which has the default conveniently set (for us, anyway) to title search. Voila! So that's a great plus.

What else is a positive about Serials Solutions' Central Search? It contains a list of our databases with descriptions. From that initial page, the user can select all or some of them to search. An advanced search screen, with pick lists for various fields, and Boolean operators, is also available. It looks similar to the advanced search screens in ProQuest or EBSCO.

When you search, a frame appears to the left showing the databases being searched and how many “hits” you are getting in each database (called “results by source”). You do not have to wait for the search to end if it is going to be a long one—you can stop it mid-search and see the results. And you can look at that left frame and decide you want to see the searches from just one of those databases—maybe the one that shows the most hits.

Sounds fantastic so far, right? Well, there are a few glitches to be resolved. For example, the system can search for ISSN, but not journal title. Have you met many students who know the ISSN of a journal they are seeking? It can limit to full text only—that's good. But the count at the top does not change after you use the filter, so it is hard to tell the difference between the total count of items found versus just the full-text items. Confusing! And the “peer-reviewed” filter has yet to work, although we keep bugging Serials Solutions about it.

This next problem concerns separate databases. I am tired of my first hits being from sources such as Saint Paul Pioneer Press. I want the standard stuff, like Newsweek, New England Journal of Medicine, or Nation. This problem might be somewhat solvable if the results could be sorted by relevancy, but at this point there is only date, author, source, and title sorting. While there are pros and cons of this system, I doubt that I could ever go back to searching databases individually. One-stop shopping is an idea whose time has come; in fact, it is overdue. Central Search has quite a way to go before it eliminates the bugs. Essentially, Central Search is in beta mode (at the time of this writing), and we are the guinea pigs. Trying to let them know what we need is frustrating and requires much work.

To give them credit, I bet that technically a lot of this stuff is hard to do, and the databases themselves are probably a hindrance when there is no uniformity. So they are trying, but meanwhile, our librarians keep constructing work-arounds instead of using Central Search as its designers had intended.

We are always waiting for technology to catch up so we can serve our users better. When are our online catalogs going to work as effectively as Amazon and our federated search engines work more like Google? At this point, I will settle for something more modest, such as the ability to search by journal title rather than ISSN.

West Virginia University's Federated Searching
Penny Pugh

Does federated searching fulfill its promise? Does it make looking for library resources just like using Google? Can students identify and locate appropriate research materials without the benefit of any prior knowledge or understanding of how information is generated and how to access it? The verdict is still out at West Virginia University Libraries (WVU Libraries). After a single semester of experience with a federated search service, I can offer some insight into the implementation of the service and some thoughts on its future at our institution.

WVU Libraries began thinking and talking about federated search services in 2002, almost by accident, as we evaluated Open URL link resolvers. We knew from user studies (LibQual and Web usability tests) that our students and faculty faced two difficulties in accessing library resources. First, the tremendous and rapid growth in our electronic collections made it difficult for them to determine if the library owned a particular title. They did not find it a simple matter to toggle between a reference in an online database and our catalog or electronic journal list to check for holdings. What seemed easy for librarians was actually confusing for our users. The second difficulty was related to the sheer number of resources available, making selection of the best database for a particular topic difficult for students. We discovered that students would return to a few databases with which they were familiar, whether or not those databases were appropriate for their research topics. At that time we considered the first problem (checking serial holdings) to be greater and sought to implement a link resolver as soon as possible. Although we were not yet in the market for a metasearch tool, vendors were anxious to include a demonstration of their federated search product along with the demo of their OpenURL software. As a result, we saw several, but we were not enticed by those early demonstrations. Therefore, we decided to address the patrons' difficulties with resource selection through a redesign of our database access pages.

In 2005, we revisited the issue of federated searching. Federated search services were becoming more prevalent in academic libraries, offering evidence of their value to researchers. Additionally, our users continued to ask for some easy way to select databases, despite our Web page redesign. We decided to take another look at federated search products. Two companies had products that appeared to meet our needs: Serials Solutions and WebFeat. After extensive testing, we selected WebFeat and began working to customize the interface and implement the service.

Implementing our federated search system took longer than we anticipated. We began customization work in May with an anticipated opening date of August, in time for the start of the academic year. We selected all the resources to be included, and library

systems staff provided WebFeat with a list of 150 databases and with access information for each. We chose to exclude some resources, including two important science resources, SciFinder Scholar and Web of Science because both have a limited number of user seats at WVU and one required the download of a software client. WebFeat then created a single Web page that offered multiple search options. The top of the page was a standard guided search form, with drop-down menus to select search fields, Boolean operators, and date ranges. The search form was followed by two options, either "search by category" or "select individual resources" from an alphabetical list. The categories were the same broad disciplinary groups that we offer on our own database access Web pages. This page was the starting point for the implementation team's customization work.

The new service had been dubbed E-ZSearch, the result of a library-naming contest, but the initial search form provided to us by WebFeat was too complicated for our taste and anything but "E-Z."5 It required users to choose a broad disciplinary group or a single database, the very choices they struggled with from our own pages. We knew that we needed to simplify, and we wanted to make our implementation more Google-like. We placed a prominent search box in the upper left-hand corner of our home page under text that reads "E-ZSearch Find Articles and Books." Searches conducted from that box include results from four sources: our library catalog, EBSCOhost Academic Search Premier, JSTOR, and Lexis Nexis. We selected these resources because we find ourselves most often recommending them to novice researchers. With the exception of our catalog, results from searches in those databases provide full-text links and they are multi-disciplinary. We easily agreed on the resources to include in the quick search category, but there was extensive discussion about whether or not to tell users up front what they were searching. Several librarians thought that we should list the databases being searched, fully informing the searchers. Another group favored simplicity and no instructions or explanation since those details existed elsewhere on our Web site. In the end, we opted for the simple search box with the direction, "find books and articles," and a link to the advanced search screen.

The advanced search interface was the place where our team developed a list of subjects and disciplines, based on the degree programs at WVU. We limited ourselves to thirty-six entries to create a screen that would not require scrolling. This required the combination of some disciplines, like agriculture and forestry. We included a category for quick search and for our digital collections. The advanced page includes a guided search form with Boolean connectors and field searching. We also included minimal instruction at the bottom of the search screen. It reads, "E-ZSearch provides a quick and easy way to search multiple databases at once. You can perform more complex searches by directly searching a database selected from our list of databases." This links to our Web pages which in turn list and describe databases.

In the interest of simplicity, we did not individually list all the resources on the E-ZSearch site, although that option was available as part of the WebFeat service. Members of the team and subject bibliographers selected the databases to include in each category, and we began with ambitious lists of resources in an attempt to help users find as much as

possible. For example, our initial list in the biology category included thirteen separate databases. We discovered quickly that federated searching of this many resources was too slow, and selectors pared their lists to five or six databases in most categories. Throughout the implementation, we needed to remind ourselves to keep it simple.

Speed was the single biggest issue throughout the development and testing of the service. In our implementation, E-ZSearch seemed to process and display search results more slowly than our pre-purchase testing of the WebFeat product. Team members felt that the service would not be acceptable to our users, who were accustomed to instant results from Web search engines, and it took our systems staff some time to diagnose and correct the problem. We discovered that the quick search box seemed slower than the advanced search service because it did not display any results until all searches were complete. With the advanced search, intermediate search results began displaying immediately, giving the impression of greater speed. Changing the quick search to operate in the same fashion helped dispel the impression of slow searching.

Two additional issues arose as we tweaked and tested the system: the order in which search results displayed and the relevancy of the results retrieved. Search results in WebFeat can be sorted by author, date, or title, ranked by relevancy, or grouped by resource. Users are able to change the sort through a drop-down menu. We initially chose the default to be grouped by resource. This choice seemed a logical and helpful way to look at results, but as we tested the system, we discovered that groups were displayed in the order in which the searches were completed. The order varied depending on the terms searched. For one search, our library catalog might be at the top of the list, for the next, JSTOR results might be listed first. We were not completely satisfied with this seeming inconsistency, but we were not really happy with the results of relevancy ranking either. No clear explanation exists of why, for example, a search on “interest rates” would list several congressional hearings from the previous century as the first results. Customer support staff at WebFeat told us that relevancy ranking displayed the first two to three results from each database searched. Of course, the native search interfaces differed in the way results displayed, so taking the first result from JSTOR might be the most relevant while the first from an EBSCOhost database would merely be the most current. Based on that information, we chose to leave “grouped by resource” as the default display.

After three months of development, the E-ZSearch service went live on the first day of spring semester in 2006. News releases went out, and articles appeared in the student newspaper and the library newsletter. Librarians discussed ways of integrating it into our information literacy and instruction programs and at our public service desks. As a first effort at evaluating the service, we gathered user satisfaction information by surveying students who enrolled in our one-credit hour library research class. We conducted the surveys on the first day of class, before any course lectures or assignments, and we asked students to test E-ZSearch and to share their overall impression, what they liked most, and what they would change. Students' overall impressions were positive. Some of the terms they used to describe E-ZSearch were: fast, quick, easy, convenient, efficient, useful, and user friendly. One student said it was “direct and easy to understand.”

Another liked it because there was “no beating around the bush.” When asked what they would like to change, student feedback indicated that they would prefer a different order in which results are displayed. They said things like “Show best matches first” and “Do a star system and rank the best books for topics.” Ironically, virtually no students indicated that the service was too slow, even though librarians were still unhappy with the slow speed. Based on our user feedback and on usability studies conducted at other institutions, we reconfigured E-ZSearch to rank results by relevancy, despite our own reservations about how relevancy is determined by WebFeat.

In the final analysis, E-ZSearch is useful but not a perfect solution to library research. The quality of the results is variable, depending on the search terms entered. Simple phrases like “global warming” or “seed germination” retrieve acceptable results, while results of many searches contain false hits.

After a semester in use, we see the need for a sustained marketing campaign. E-ZSearch is prominently displayed on our home page and should be getting much more usage. A database such as Academic Search Premier, for example, must be located by looking through a set of menus, and yet the use of the database is significantly greater than E-ZSearch. Our statistics show that 30,343 searches were conducted in E-ZSearch during the first semester while, as a point of comparison, more than 236,000 searches were conducted in Academic Search Premier during the same period.

Finally, we must resolve concerns about federated searching among our own library faculty. There is a discernable tension from the dilemma of wanting to improve information literacy skills among our students while providing simple one-stop shopping via E-ZSearch. Though we worked diligently to develop the E-ZSearch interface, we are careful not to promote the service as a replacement for thorough research. We remind users that E-ZSearch is a good starting place for research and best serves novice researchers seeking a few good sources. But, as good librarians, we are far from conceding to the “good enough” approach to research.

Exploring the Federated Search at Northwestern University Library
Frank Cervone

Research has always been a challenging endeavor, but there are so many resources available online today that finding a much-needed article can feel like an insurmountable task. It was with this task in mind that the Northwestern University Library embarked on a search for a federated search tool.

In the spring of 2002, at a time when Google Scholar and Microsoft Academic Live were just vague concepts in some strategist's head, we embarked on a system selection process for a federated search system. In 2002, there were four major vendors of federated search products: WebFeat, Fretwell-Downing, Endeavor Information Systems, and Ex Libris.

Since WebFeat was the first product to market, we talked to them first. However, at that time, WebFeat did not use a standard protocol, such as Z39.50, for harvesting results.

Instead it issued a search through the user interface of the database and then “screen scraped” the results, programmatically processing the resulting screen output to generate a result set. Given the volatility of search result output display and the consequent need for continuous maintenance to address these changes, we did not feel this was a good, long-term strategy; moreover the steadfast refusal of WebFeat to provide a combined, deduplicated results list from multiple searches severely limited the functionality of the product, in our opinion.

By then, our library management system vendor Endeavor Information Systems had released Encompass. Consequently, we felt there was an advantage in considering Encompass. Although we were impressed with the ease in which the user interface could be customized, Encompass could not deliver the level of performance we needed.

In the spring of 2003, we began looking at products from Fretwell-Downing and Ex Libris. Some ARL (Association of Research Libraries) members were working with Fretwell-Downing on the “Scholar's Portal” project, while several others were working with Ex Libris. After extensive evaluation of both vendors, we chose Ex Libris' products as they leveraged existing technology, such as Z39.50, but also provided for future extensibility through XML gateways and newer protocols such as SRU/SRW. In contrast, Fretwell-Downing was exclusively focused on Z39.50 connectivity at the time and did not offer an OpenURL linking product.

Given that background, we began implementing Ex Libris' SFX® (OpenURL linking) and MetaLib®, their federated search product. While the two services are often sold as an integrated whole, typically the implementation processes are separate with OpenURL linking usually being installed first. Implementing SFX was quicker than that of MetaLib. Where SFX took about three months to implement, MetaLib took almost a year6.6 The difference in implementation time is primarily due to the increased complexity of federated search. In our experience, this complexity was primarily related to four issues: Z39.50 connectivity, resource classification, database performance, and user interface customization.

As we were to discover, ongoing maintenance- of Z39.50 connection information can be quite labor- and knowledge-intensive. While Ex Libris provides a knowledge base of Z39.50 connection information, this information must be customized for the local environment. For example, the local site must define the URL used to access the database as well as the user id and password used for the Z39.50 connection. Additionally, someone must select what is included in the various database subscriptions. Presumably, configuration of databases would be a relatively static environment, but the reality is quite different, especially for aggregated databases where configurations change frequently.

In many federated search system installations, an especially difficult aspect of implementation is the categorization of resources. Before databases can be defined in the system, a scheme for subject classification must be in place. At Northwestern we were able to avoid many of the problems by streamlining the process using our university

catalog as a guide. By defining subject areas in direct relationship to the departments, schools, and institutes of the university, a small group of bibliographers and reference staff was able to come up with a first pass of suggested subject classification for each database. Consequently, the subject specialists did not have to generate their own lists from scratch, and we greatly reduced our implementation time.

Another thorny problem relates to selecting databases for "QuickSearch" groups (groups of preselected databases that are searched together). We need to investigate whether the databases within each group all support the same types of searches (author, title, keyword, etc.) and if they return results in the same manner (ascending by year and author, relevance ranked, etc.). Most of the time, the answer is no, which means that we must tweak the Z39.50 configurations, if possible, to make them perform more consistently.

In addition, the performance characteristics of databases within a group must be tested. Our research has found that some databases always respond more slowly than others. While this may not be a noticeable problem when searching a single database, the response time differences become quite obvious when searching within a group of databases. In the end, tough choices may have to be made. In general, we will not include certain databases that always take too long to respond (approximately more than fifteen seconds after all the others in the group) in "QuickSearch" groups. If possible, we substitute another database that is similar in content or try to find a different vendor at contract renewal time.

With experience, interface issues have also become evident. Although the point of federated searching is to make searching as simple as possible, the model is unfamiliar to most people since it does not work like Google, which is what people expect. Google does not ask you to select a database or subject area. Not surprisingly, usability research at Northwestern and other institutions has found that people clearly prefer a simple search because it uses a default set of databases and avoids complex query statements, such as Boolean operators.7

Another problem area for most users is related to combining databases in unique ways outside of the "QuickSearch" groups. In our usability studies, as well as others, those conducted by users continue to find this confusing and try to avoid it if possible. Unfortunately, the MetaLib interface is not readily customizable. This has led many institutions either to minimize the amount of customization (to the detriment of usability) or to develop their own front end to the software.

Of these four issues, only one is directly related to choice of software. Both resource classification and database performance are issues that apply to any federated search implementation regardless of vendor. Z39.50 connectivity issues can be avoided by choosing a vendor that hosts the federated search product. However, hosted products are less customizable. This is particularly true in relationship to the user interface.

Ideally, the user interface of a federated search product should be seamlessly integrated into the information-seeking environment of the user. However, most products still have a long way to go in this regard. For most hosted services, the only customization to the user interface that is possible is to add the logo and identifying information of the institution. Locally hosted systems provide options for customization that are not possible with a vendor-hosted system, but unless the user interface is based on newer standards, such as XSL/XSLT, customization can be difficult and cumbersome.

If federated search is to make greater inroads, we must quickly come up with better tools to integrate these services into the user's information-seeking environment, be that the library Web site, a course management system, or some other information portal of the patron. If not, other services such as Google Scholar and Microsoft Academic Live will quickly fill the void and another opportunity for libraries will have been lost.

References

1. The Pitt Digital Library Home Page at the University of Pittsburgh is available at http://www.library.pitt.edu (accessed July 30, 2006).

2. An Example of the Zoom Advanced Search Interface is Available at http://0-search3.webfeat.org.wncln.wncln.org:80/UPittSearch.asp?int=a (accessed July 30, 2006).

3. Roy Tennant, “Cross-Database Search: One-Stop Shopping”, Library Journal 17 (October 15, 2001) (126), pp. 29–30.

4. The Grossmont College Library's Home Page with Central Search Listed under Databases is available at http://www.grossmont.edu/library/ (accessed July 30, 2006).

5. The West Virginia University Library Home Page with E-ZSearch Displayed is available at http://www.libraries.wvu.edu/ (accessed July 30, 2006).

6. The Northwestern University Library Home Page with their Adaptation of MetaLib is available at http://www.library.northwestern.edu/ (accessed July 30, 2006).

7. Frank Cervone, “What We've Learned From Doing Usability Testing on OpenURL Resolvers and Federated Search Engines”, Computers in Libraries 25 (October 2005) (9), pp. 10–14.