

# Key issue

## Federated searching: today, tomorrow and the future (?)



**FRANK CERVONE**

Assistant University Librarian  
for Information Technology  
Northwestern University,  
Evanston, IL, USA

Although for many people it may seem that federated search software has been around forever, it has only been a little over three years since federated search burst on the library scene. In that time, the software has been evolving in many ways as the landscape of Internet-based searching has shifted.

When many libraries decided to install federated search software, they were trying to address some of the issues that have confounded searchers since the introduction of CD ROM-based indexes in the early 1990s. For example, it has been known for some time that when patrons have multiple databases available to them, they typically prefer to use a single interface rather than having to learn difference interfaces for each database. Furthermore, most people want to have a way to search those multiple resources simultaneously and then consolidate the search results into a single, ordered list. More or less, these problems are now solved through the use of federated search tools.

But federated search tools can also provide other benefits to patrons. Many search tools are able to hook into local authentication systems. By doing so, access to resources can be controlled, making it possible to only expose those resources to which a patron can actually gain access. In large, complex environments where a database may not be purchased for organization-wide use, this is a critical management issue. A side benefit of this authentication linkage is that it also provides a convenient mechanism for enabling remote

authorization, thereby allowing patrons the ability to use the resources from wherever they may happen to be.

### Standard features and evolving trends

In the current generation of federated searching products, standard features include the ability to use multiple protocols for gathering data from information products. In addition to the ubiquitous Z39.50 protocol, newer protocols such as SRU/SRW and OAI-PMH are rapidly becoming basic requirements for all federated search engines, regardless of scale.

As the needs of researchers and faculty evolve, advanced features are being built into next-generation federated searching products, such as integration of the federated searching software with other packages like course management systems (Blackboard, WebCT, and SAKAI), bibliographic management software (RefWorks and WriteNote), as well as campus portal software. Another emerging trend is expanded functionality, such as tighter integration of the catalog with full-text sources and mechanisms for displaying results through visualizations. These visualizations may take the form of a lattice diagram that links keywords together (Aquabrowser), a concentric diagram that clusters similar items (Grokker), a 'phrase cloud' that takes subject headings and other information and displays it in a tag cloud

format (Encore), or a faceted breakdown of the structured heading into a more usable form (Primo and Endeca).

### **A quick survey of the current marketplace**

As is true in other software markets, the library management software marketplace has consolidated even further with the recent merger of Endeavor Information Systems into the Ex Libris Group. Unlike the Sirsi/Dynix merger, where the two product lines were continued, the Endeavor products (ENCompass and Discovery: Finder) will be going by the wayside in favor of Ex Libris' MetaLib. However, the focus of MetaLib itself is changing as well. In the future, while MetaLib will still continue to have its own native interface, expect to see more libraries using Primo as the front-end to both MetaLib and their Voyager or Aleph catalogs.

In the short term, minor increases in functionality in existing federated search products will be the norm as the vendors put the finishing touches on their next-generation integrated interface solutions. An example of this short-term approach is that both Ex Libris (in MetaLib) and Serials Solutions (CentralSearch) are integrating the Vivisimo faceted browser into their federated search interfaces.

Further evidence of the maturing of the marketplace for federated searching is the introduction of lighter-weight solutions for smaller libraries. Most of these solutions come from traditional content aggregation services. Seeing the writing on the wall, these vendors are trying to carve out a new space, which has led to the launch of hosted federated search services from Cambridge Scientific Abstracts (CSA) through their Illumina platform, the integration of federated search and Grokker visualizations into the EBSCO database platform, as well as the integration of the MuseGlobal search engine platform into Ovid. Additionally, several vendors (Serials Solutions and Ex Libris) have moved into the application service provider space (ASP) by providing hosted services for their federated search products. WebFeat has taken this a step further with the introduction of *WebFeat Express*, a hosted solution that is configurable and ready to use by simply answering some basic questions about the library and its resources.

### **Some things to bear in mind**

With the widespread use of federated search, a number of things have been learned about how people actually use these products. For example, multiple usability studies report that one of the distinct advantages for people using a federated search engine is the 'serendipity' of finding previously hidden information resources. Since most researchers are only familiar with a limited number of databases in their area of expertise, they rarely search in databases outside their immediate discipline. The use of a federated search engine frequently will take them into new databases, thereby providing new worlds of resources to explore. The end result is that students and faculty often will find many significant citations of which they were not previously aware.

Usability studies also point out the need for simple search interfaces. Patrons, in general, do not want search interfaces that provide a myriad of choices or require them to make decisions. Furthermore, the vast majority of patrons will never use a 'complex' or 'advanced' query option as they expect the search engine to figure out what it is they want. The reality is that most of these complex or advanced options are so complex or advanced that the average person cannot really work out what it is they are supposed to do, so the functionality goes unused.

Consequently, creating an effective federated search interface is a delicate balancing act. Perhaps one of the most important aspects in creating a usable interface is finding the right grouping of databases for subject areas or 'interests.' Defining these is typically the most complex and time-consuming aspect of the implementation. If we go overboard, however, we will run into the problem of 'too much leads to too little.' Long lists of resources are often developed in the hope of promoting awareness of the various resources in the library; however, students find these lists confusing. As one student in a recent study put it: "Those long lists really make me feel stupid." Obviously, a product that makes someone feel stupid is not one to which they will return. Best practices indicate that one way of alleviating this problem is to have a 'best bets' group of 3 primary databases in a subject area, with the rest of the relevant resources as a secondary group afterwards.

### Summary

As the federated search market moves ahead, one of the main things we are seeing is the integration of more advanced search technologies, both on the back end and the front end. This provides the software with the ability to better interface with other commonly used software in the research environment, such as course and bibliographic management systems. Combined with the incorporation of visualization and topic clustering tools, these products are meeting the emergent needs of researchers.

However, all is not rosy. Many people question the long-term viability of federated search software in a world where Google Scholar and Microsoft Academic Live are able to search much, although not all, of the same content that federated search engines do. A big question for both vendors and libraries is where federated search will fit into the search environment of the future. Will Google Scholar and MS Academic Live completely eliminate the need for federated search? Or will the next generation of federated search engines offer extra value to the searcher in the form of highly contextualized information in the local environment? The answer to these questions remains to be seen, but it is clear that the overarching concept of an integrated searching of resources is here to stay.

### Further sources of information

One of the earliest articles related to issues in federated searching:

Payette, S and Rieger, O, Z39.50: The User's Perspective, *D-Lib Magazine*, April 1997. Online at: <http://www.dlib.org/dlib/april97/cornell/04payette.html>

A recent discussion of the tensions between federated search engine implementations and the ubiquity of the Google search engine:

Chen, X, Metalib, WebFeat, and Google: The strengths and weaknesses of federated search engines compared with Google. *Online Information Review* 2006, (30)4, 413–427.

A good overview of SRU/SRW and OAI-PMH:  
Fox, R, Lingua franca of digital libraries. *OCLC Systems & Services: International Digital Library Perspectives*, 2006, 22(1), 26–33.

### Representative Vendor listing:

- AGen<sup>TM</sup> – Auto-Graphics, Inc:  
<http://www.auto-graphics.com/>
- Aquabrowser – Medialab Solutions BV:  
<http://www.aquabrowser.com/>
- CARLweb – TLC, The Library Corporation:  
<http://www.tlcdelivers.com/tlc/carlweb.asp>
- CentralSearch – Serials Solutions:  
<http://www.serialssolutions.com/promotion/centralsearch.asp>
- Chameleon iPortal – VTLs Inc:  
<http://www.vtls.com/Products/gateway/>
- DQM2, Deep Query Manager<sup>TM</sup> and Deep Federation Portal<sup>TM</sup> – BrightPlanet<sup>®</sup> Corp:  
<http://www.brightplanet.com/products/dqm.asp>
- EBSCO AtoZ with WebFeat – EBSCO Information Services:  
<http://www.ebsco.com/home/ejournals/default.asp>
- Encore – Innovative Interfaces Inc:  
<http://www.iii.com/>
- Explorit<sup>TM</sup> Software Suite – Deep Web Technologies LLC:  
<http://www.deepwebtech.com/technology/explorit.php>
- FDI Portal – Fretwell-Downing:  
[www.fdgroupp.com/fdi/marketing/zportal\\_choice.html](http://www.fdgroupp.com/fdi/marketing/zportal_choice.html)
- Grokker – Groxis, Inc:  
<http://www.grokker.com/>
- Horizon Information Portal – SirsiDynix<sup>TM</sup>:  
<http://www.sirsidynix.com/Solutions/Products/portalsearch.php>
- Keystone<sup>TM</sup> DSL – Index Data ApS:  
<http://www.indexdata.dk/keystone/>
- MetaLib – Ex Libris Group:  
<http://www.exlibrisgroup.com/metalib.htm>
- MetaSearch Solution – Blue Angel Technologies, Inc:  
<http://www.blueangeltech.com/>
- MultiSearch – CSA:  
[http://www.csa.com/e\\_products/MS\\_main.php](http://www.csa.com/e_products/MS_main.php)
- Muse Solutions and MuseSearch<sup>TM</sup> – MuseGlobal, Inc:  
<http://www.museglobal.com/solutions/>
- OpenSiteSearch – Hosted by SourceForge<sup>TM</sup>.net:  
<http://opensitesearch.sourceforge.net/>
- Polaris<sup>®</sup> PowerPAC<sup>TM</sup> Portal – GIS Information Systems, Inc.<sup>TM</sup>:  
<http://www.gisinfosystems.com/>

- Searcher Analyzer – TDNet Inc:  
<http://www.tdnet.com/site/page.asp?ID=460A&Parent=457>
- SearchSolver – Ovid, Inc:  
<http://www.ovid.com/site/products/tools/searchsolver.jsp>
- WebFeat Prism and WebFeat Express – WebFeat Inc:  
<http://www.webfeat.org/>
- ZONE-Pro™ and Zones™ line of products – BiblioMondo, Inc:  
<http://www.bibliomondo.com/site/solutions/search.php>

Article © Frank Cervone

To view the original copy of this key issue, published in *Serials*, the journal of the UKSG, click here:

<http://serials.uksg.org/openurl.asp?genre=article&issn=0953-0460&volume=20&issue=1&spage=67>

For a link to the table of contents for the issue of *Serials* in which this article first appeared, click here:

<http://serials.uksg.org/openurl.asp?genre=issue&issn=0953-0460&volume=20&issue=1>