



# Smart Libraries™

Formerly Library Systems Newsletter™

50 East Huron Street, Chicago, Illinois 60611-2795, USA

July 2004 Volume XXIV Number 7

## Elsevier takes on Google and ISI

Elsevier is gearing up to compete with ISI with the release this fall of its huge abstract and indexing (A&I) service, Scopus. As the dominant global publisher in science, Elsevier receives a lot of negative publicity on its pricing policies and positive publicity on its information products.

Two years ago it launched Scirus, which is gaining scientist's attention as it screens out the noise generated in searching scientific topics by eliminating irrelevant hits that are presented by traditional Web search engines such as Google.

Google has been actively recruiting academic content by agreeing to index MIT's institutional repository and scholarly society publications, such as those in databases available from the Institute of Electrical and Electronic Engineers (IEEE) and the Association of Computing Machinists (ACM).

### Scopus debuts this fall

Scopus indexes 80% of peer-reviewed scientific literature and covers 13,000 international journals from more

See **Elsevier** on page 2

## XML to the rescue for metasearch

A year ago librarians characterized metasearch technology as immature and the projects that used it as prototypical or experimental. Metasearch has turned the corner in effectively searching the ever-growing collections of electronic resources.

The technology has gained a higher level of sophistication and maturity. Many library organizations have rolled out successful implementations, and metasearch product sales are flourishing.

Significant work remains to be done though. Much of the technology continues to be based on *ad hoc* methods versus a set of open standards. Building a metasearch environment that fully supports all of a library's licensed resources is still a major challenge.

A key problem in developing interfaces for searching multiple electronic resources simultaneously involves the use of behind-the-scenes search and retrieval protocols—the mechanism to send commands and receive results.

Z39.50 and SRW (Search and Retrieve Web) provide the means for search and retrieval using open standards, but these standards are not necessarily widely

See **XML** on page 4

## IN THIS ISSUE

Elsevier Takes on Google and ISI  
PAGE 1

XML to the Rescue for Metasearch  
PAGE 1

Librarians Pilot Support of Online Learning  
PAGE 2

Xobis—The Future of MARC  
PAGE 3

Move Over Google—Context Matters  
PAGE 5

CDL Selects MetaLib for Metasearch Project  
PAGE 5

Metasearch Initiative Gaining Steam  
PAGE 6

Automation Sales Sluggish  
PAGE 6

AGent Powers JerseyClicks  
PAGE 7

Library of Congress Chooses SFX  
PAGE 7



### Receive *Smart Libraries* via e-mail

Subscribers who would like an e-mailed version of the newsletter each month should forward their e-mail address and ALA identifier (the 7-digit number printed on the top line of the address label that appears on page 8 of your newsletter) to [jfoley@ala.org](mailto:jfoley@ala.org). Type "e-mail my Smart Libraries" into the subject line. Issues will be e-mailed in addition to your print subscription and at no additional charge.

Elsevier *from page 1*

than 4,000 scientific, technical, and medical (STM) publishers. Many titles are indexed since 1999.

In development for the last two years with 20 universities, Scopus is completing beta trials now with Pepperdine University in California, University of Toronto in Canada, and Oxford University in the United Kingdom among others. These schools' faculty have been providing feedback on usability aspects.

Described as the biggest A&I database of scientific literature ever assembled, Scopus simultaneously searches the scientific Web using Scirus, its science-only Internet search engine. Data from Embase (medical A&I) and Compendex (engineering A&I), also owned by Elsevier, are expected to be included in addition to the 1,800 journals that appear in Elsevier's Science Direct.

The system is designed to work the way researchers do. Scopus developers have conducted considerable research on scientists' cognitive patterns in using abstract and citation data. The user

interface has intuitive browsing tools, and users can search straight from the homepage, linking from the search results page to the full text. Primary benefits for scientists, researchers, and students are that Scopus:

- Provides simple searching
- Eliminates duplicate search results
- Ensures seamless access by eliminating dead links

Designed for the user and developed for the librarian, Scopus includes Counter-compliant usage reports tailored to customer needs and offers on-site training.

### Scirus wins recognition

Elsevier partnered with Fast Search and Transfer ASA (Fast), a leading developer of search, filter, and compression technology

## LIBRARIANS PILOT SUPPORT OF ONLINE LEARNING

OCLC has completed the first phase of an e-learning pilot that combines the flexibility of virtual reference with the infrastructure of course management software. By providing course support online, librarians are available to their users without requiring students to visit the library or attend an instruction session.

Eight institutions experimented by linking the homepage of selected online courses to enable students to readily access OCLC's QuestionPoint. Participating institutions included: Baylor University, City University Online, College of Charleston, Manhattanville College, Sage Colleges, University of Illinois at Chicago, University of Nevada at Las Vegas, and University of Edinburgh.

Students liked the system so much they began using QuestionPoint in courses that were not yet linked. This service connection integrates the library into the curriculum in a way that reaches the students at their point of need.

An analysis of the forms that track usage data and feedback during the pilot will be used in planning to expand support for more courses.

QuestionPoint is a virtual reference service that offers an infrastructure of software tools and communications that are supported by a global network of libraries. It enables delivery of the service and draws on the community to share its expertise in providing reference service online.—*JL*

Contact: [www.oclc.org/questionpoint](http://www.oclc.org/questionpoint)



to create Scirus, so scientists could identify structured and unstructured information that is both publicly available on the Web and in proprietary scientific databases.

Scirus, named for a Greek seer, is a free tool that identifies free websites and fee-based documents. It employs linguistic algorithms in searching 167 million scientific documents on the Web, including full-text journal articles, preprints, academic studies, and scholar's homepages. Scientists, researchers, and students can rely on Scirus to:

- Filter out nonscientific sites
- Find peer-reviewed articles that aren't visible to other search engines
- Search within selected subject areas including health, life, physical, and social sciences
- Narrow the search to a particular author, journal, or article

- Restrict results to a date range
- Find conferences, abstracts, and patents
- Refine, customize, and save searches

Users can choose to rank their search results by relevance or by date. Relevance is determined by two factors: the location and frequency of a search term within a result, and the number of links to a page.

Since it was launched three years ago, Scirus was voted the Best Specialty Search Engine by the Editors of SearchEngineWatch.com two years in a row (2001 and 2002) and was nominated Best Science Web Site for Webby Awards in 2003 by the International Academy of Digital Arts and Sciences.—*Judy Luther*

Contact: [www.news.scopus.com](http://www.news.scopus.com)  
[www.scirus.com](http://www.scirus.com)

# XOBIS

## the future of MARC

Dick Miller, head of Technical Services at the Lane Medical Library at Stanford University Medical Center, has been hailed as a visionary for leading the effort to develop Xobis, the XML Organic Bibliographic Information Schema. Xobis enables libraries to create expanded records using MARC data elements in a format that allows them to be integrated with other Web resources.

Although the Library of Congress has created MARC XML, which is a literal translation of MARC, it does not address the challenges of finding electronic documents in a Web environment. Current standards (AACR and MARC) focus on describing a work rather than providing data useful to access it.

Miller distinguishes between descriptive metadata, which identifies characteristics of an item, and managed metadata, which identifies the location of an item. Going beyond the flat record structure of MARC, Xobis uses XML (eXtensible Markup Language) to define the relationship between data elements. It combines bibliographic and authority data into a

single structure that is less complex than MARC, more complete than Dublin Core, and offers the flexibility of XML.

Libraries are starting to use XML to describe features of digital resources that enhance interoperability. New structures are needed that allow for indexing full-text objects. Xobis emphasizes interoperability first, followed by a detailed representation of objects.

Xobis is one of three initiatives from the Medlane project, which was started in 1998 as an experiment to explore how XML could improve library services. Xobis is viewed as building on prior standards work within the industry and was presented to Functional Requirements for Bibliographic Records (FRBR) Working Group at International Federation of Library Associations (IFLA) last summer. Additional information and comments are welcome.—*JL*

Contact: <http://laneweb.stanford.edu:2380/wiki/medlane/schema>

XML from page 1

supported beyond library OPACs and may not be the optimal choice for full-text information resources.

## Single search protocol needed

Today's environment of information resources lack a single search protocol supported among all the products and services. In the absence of a protocol shared by a metasearch interface and a given information resource, many vendors have developed their own specialized gateways that exchange search and retrieval data formatted in XML (eXtensible Markup Language).

All the major commercial developers of metasearch interfaces, such as MuseGlobal, Fretwell-Downing, Inc.; Ex Libris (USA), Inc.; WebFeat; and Endeavor Information Systems Inc., include XML gateways among their techniques for communicating with electronic content sources. The XML gateway provides a more efficient approach for integration into a metasearch environment than some alternate approaches.

One metasearch technique often employed for searching resources that do not support Z39.50 or other search and retrieval protocol involves intercepting HTML and parsing it to extract search results. This method operates inefficiently and is fragile. Anytime the information resource changes the way it presents information through its Web interface, the parsers have to be updated.

## Libraries create gateways

Until now, no tools were available to help a library make the electronic resources it creates interoperable in a metasearch environment other than by adding Z39.59 support.

Libraries that use Endeavor's ENCompass, beginning with Version 3.5, can develop their own XML gateways. Endeavor has developed a toolkit for creating an XML gateway between ENCompass and other digital collections or electronic resources. This toolkit, called the ENCompass Outbound XML Search API, allows the provider of an information resource and an institution licensing ENCompass to work together to make that resource searchable by ENCompass users.

An API, or applications programming interface, provides a set of technical specifications that can be used by programmers to extend the functionality of a system. In this case, the API provides a framework for communicating with a site that uses ENCompass for exchanging search and retrieval information when no other convenient mechanism, such as Z39.50 or SRW, is available.

Endeavor especially recommends using an XML gateway for resources not based on MARC records, since XML offers better flexibility in searching nonbibliographic data. This API allows libraries to engineer their own local information resources and digital collections to interoperate with ENCompass.

Although the ENCompass Outbound XML Search API allows libraries that license the product to create XML gateways for their local digital collections and other electronic resources that might otherwise not be supported well by ENCompass, Endeavor uses this approach with the major publishers of scholarly content on behalf of its entire customer base of libraries that license ENCompass.

The company, for example, developed an XML gateway between ENCompass and Ebsco resources. These resources were previously searchable through ENCompass using the Z39.50 search and retrieval protocol; however, the XML gateway technology provides better search and retrieval features and performance.

Endeavor promotes the use of the XML gateway approach over Z39.50 for nonbibliographic information such as full-text electronic journal content. It uses XML gateways in ENCompass to search other resources such as PubMed, ScienceDirect, and LexisNexis.

Although the XML gateway approach stands as the state-of-the-art in behind-the-scenes support for metasearch functionality, it stands in an unsteady state. Each vendor—both on the information provider side as well as the technology side—defines its own approach for implementing it.

Metasearching will be greatly improved once a uniform framework is defined and adopted as a standard. Only then will XML truly rescue the metasearch problem.—*Marshall Breeding*



# Move over Google—context matters

Google is scheduled to launch its IPO, and competition is gathering on the horizon. Competitors focus on either personalizing results, presenting results visually, or generating ad revenue. Each takes a different approach to determining the context for a user's search, which is then used as a filter in presenting search results and subsequent options.

Both Vivisimo (used by HighWire) and Groxis (demoed in the Dynix booth at the American Library Association show in Toronto) present results visually grouped by context (that is, Paris Hilton—the person versus the hotel). Two new ad revenue focused search engines, Quigo and Industry Brains, are generating ad revenue based on click-throughs.

Those offering personalization have names as strange as Google's, such as Dipsie, Mooter, and Eureka:

- **Dipsie.** Founded in 2003, Dipsie (Chicago) crawls the hidden Web and organizes its results based on the way people use language. In deep sea fish-

ing, a dipsie sinks a fishing line, taking it well below the surface. This description characterizes the dipsie.bot's ability to search and subsequently return users to pages hidden behind cookies, templates, and databases that foil attempts by normal methods.

- **Mooter.** A legal term, based on a moot point or one that is open to debate, Mooter (Australia) is focused on decoding Web page themes and users' implicit search patterns to deliver personalized results. Its founder brings a psychology background to addressing search questions and Mooter uses linguistic analysis to group websites for better search results.

- **Eureka.** Search results in Eureka (San Francisco) are personalized based on preferences of friends and colleagues in one's social network. This methodology has tremendous implications for employees of a company, segments of an industry, or even librarians as a profession. Searches of others in the selected network are displayed on the right side

of the screen, and the user can view lists of both recent and popular searches and sites for those within their network. Searches are shared (unless marked private) and results are influenced by filters that fit different communities. Knowing what others are searching for allows another level of communication. Names are not affiliated with searches and the larger the network involved, the more anonymous the searcher becomes. Eureka doesn't replace Google but offers another layer of filters that produce more effective search results.

Keeping an eye on these new players gives librarians a sense of users' expectations. Users are beginning to expect the systems they use to learn about them, saving them time and providing a higher level of service.—*JL*

Contact: [www.dipsie.com](http://www.dipsie.com)  
[www.mooter.com](http://www.mooter.com)  
[www.eureka.com](http://www.eureka.com)  
[www.quigo.com](http://www.quigo.com)  
[www.groxis.com](http://www.groxis.com)  
[www.vivisimo.com](http://www.vivisimo.com)

## CDL selects MetaLib for metasearch project

The California Digital Library (CDL), the virtual library that serves the University of California (UC) system and a leader in digital library programs, selected MetaLib from Ex Libris (USA), Inc., as the underlying technology platform for its Metasearch Infrastructure Project.

The UC system plans many metasearch portals to support specific searching needs. One of the early projects will be the creation of a search interface targeted toward undergraduates

for finding information on general topics. Specialized discipline-specific interfaces also are planned.

In September 2003, the National Science Foundation awarded CDL a two-year grant to extend and enhance the National Science Digital Library through integrating this resource with other science collections and resources. MetaLib will be the technology platform used for this project.—*MB*

## Metasearch initiative gaining steam

In March 2004, SLN reported a new NISO metasearching initiative. Metasearch, also known as federated searching, has led to a burgeoning industry emerging from middleware vendors, publishers, and the integrated library system (ILS) industry.

But much of the business is based on the cottage industry of HTML parsing and screen-scraping to present data from a proprietary Web interface in a single search result screen. To solve that problem, NISO's Metasearch Initiative, which now has more than 60 participants from five countries, held its first meeting in North Carolina's Research Triangle Park April 22-23, 2004.

The group's libraries, vendors, and publishers have already defined problems, created glossaries of terms, and performed environmental scans of the existing marketplace. The group is moving ahead quickly with three subgroups: access management, collection description, and search and

retrieval. Whether the group will endeavor to create a new standard or endorse existing ones has not been determined, but arriving at a set of best practices will certainly be a key objective.—AKP

**Contact:** Working documents and committee deliverables are available on the group's wiki: [www.lib.ncsu.edu/niso-mi](http://www.lib.ncsu.edu/niso-mi).



## AUTOMATION SALES SLUGGISH

2003 proved to be a relatively slow year for integrated library system (ILS) vendors. A review of the sales by library automation companies April through May 2004, shows sales continue to be sluggish, with a moderate number of contracts, mostly to smaller libraries.

The only contrast in any library technology sales was in the first quarter of 2004 when Ex Libris reported sales of MetaLib with SFX to more than 40 university campuses in the United States and abroad. This single vendor is capturing the OpenURL linking arena, compared with the more competitive nature of other product markets.

Sirsi Corp. reports that the University of Puerto Rico has licensed the Unicorn library management system for its Mayagüez campus of 13,000 students. The Ocean County Community College in New Jersey licensed Unicorn to manage its collection of over 100,000 items. The Edwin A. Bemis Public Library in Littleton, Colo., will implement the ASP version of Unicorn, migrating from a Carl system. The U.S. Navy Department Library selected Unicorn to replace a Library.Solution system.

Second quarter sales announced by Dynix for its Horizon system include Siena Heights University Library in Adrian, Mich., migrating from Dynix Classic; Holy Names College in Oakland, Calif., for the first-time automation of its library of more than 100,000 items.

Sales have been strong for Dynix in Canada in the last two months. Three small Canadian libraries selected Horizon to replace their Dynix systems. The Saskatchewan Provincial Library, with a collection of more than 1.7 million items, also will replace its Dynix system with Horizon. In Ontario, the Cobourg Public Library and the Port Hope Public Library both selected Horizon to replace their Sirsi MultiLIS systems.

In Scotland, the 15 public library branches in Stirling are moving from Dynix to Horizon. The Murray Public Library in Utah also selected Horizon to replace its Dynix system. St. Marks, a private college prep school in Dallas will implement the ASP version of Horizon to replace its Dynix system.

Innovative Interfaces, Inc., licensed its Millennium system to the University of Namibia. More than 300 Innovative installations are

## Auto-Graphics

## ExLibris

## AGent powers JerseyClicks

Public library patrons throughout New Jersey can now search high-quality electronic resources through a metasearch portal created using the AGent portal technology developed by Auto-Graphics, Inc. Dubbed JerseyClicks, New Jersey's One Stop Library Gateway to Quality Information, the system was a collaboration of the Infolink Eastern New Jersey Regional Library Cooperative, the New Jersey State Library, and Auto-Graphics.

JerseyClicks simultaneously searches the full text of magazine and journal articles and online encyclopedias. It was funded by the New Jersey State Library, the state's Regional Library Cooperatives, and a federal Library Services and Technology Act (LSTA) grant.—*MB*

## Library of Congress chooses SFX

Any technology decision made by the Library of Congress (LC), the world's largest library, is noteworthy. LC has licensed the SFX, the OpenURL-based linking product from Ex Libris (USA), Inc., in a competitive federal procurement process.

This year's procurement of SFX is consistent with the trend of LC to license existing products rather than build its own systems, as it had done in its earlier automation history.

LC implemented the Voyager library management system in 1999 to replace its numerous custom-developed automation systems and has continued to acquire commercial software for its other automation needs. In 2003 it licensed the Electronic Resource Management module from Innovative Interfaces, Inc.—*MB*

outside the United States in 40 countries. The Università Commerciale Luigi Bocconi, a business school in Milan, Italy, selected Millennium to replace an Ex Libris Aleph 300 system. The King County Library System, with 43 branches serving the Seattle area, selected Innovative's Millennium system to replace its Dynix Classic system.

Endeavor licensed Voyager to Santa Rosa Junior College in California to replace a Dynix Classic system and to Texas Southern University, with its collection of more than 450,000 volumes, to replace its Geac Advance software.

Sales for Aleph 500 from Ex Libris in April and May 2004 included Lawrence University in Appelton, Wis., replacing a DRA Classic system, the library of the Italian Senate, and Liverpool John Moores University in the United Kingdom, replacing Dynix classic.

Montreal-based BiblioMondo licensed its Portfolio ILS and its Zones library portal and metasearch environment to Paris for its Paris Special Libraries Network. Three hundred workstations at

11 locations in Paris access the special collections totaling 5.2 million items in disciplines of history, law, public policy, fine arts, film, and tourism.

The Library Corp. continues its drumbeat of sales of Library.Solution to small and mid-sized libraries in the United States. Recent sales include the Mary Wood Weldon Memorial Public Library in Glasgow, Ky.; Parkland Community Library in Allentown, Pa.; Rahway Free Public Library in Rahway, N.J.; Knox County Public Library based in Barbourville, Ky., and Lexington Public Library in Kentucky.

GIS Information Systems reports that the Auburn Public Library and Talladega College Library in Alabama, Switzerland County Public Library, and the Lawrenceburg Public Library District in Indiana will all be migrating from Galaxy to Polaris. The Halifax County Library System in North Carolina will be moving to Polaris from a Follett system.—*MB*





**ALA TechSource**  
www.techsource.ala.org

Smart Libraries Newsletter  
American Library Association  
50 East Huron Street  
Chicago, IL 60611-2795 USA

---

NON PROFIT  
US POSTAGE  
PAID  
PERMIT 1479  
ROCHESTER, NY

---

## July 2004 Science A&I competition mushrooms

### Smart Libraries Newsletter

Smart Libraries Newsletter delivers hard data and innovative insights about the world of library technology, every month.

**Contributing Editors**  
Marshall Breeding  
615-343-6094  
marshall@breeding.com

Priscilla L. Caplan  
352-392-9020, ext. 324  
pcaplan@ufl.edu

Judy Luther  
610-645-7546  
judy.luther@informedstrategies.com

Andrew K. Pace  
919-515-3087  
apace@unity.ncsu.edu

**Editor**  
Chris Santilli  
630-495-9863  
chris@wordcrafting.com

**Administrative Assistant**  
Judy Foley  
800-545-2433, ext. 4272  
312-280-4272  
jfoley@ala.org

### TO SUBSCRIBE

To reserve your subscription, contact the Customer Service Center at **800-545-2433, press 5 for assistance**, or visit **www.techsource.ala.org**.

The 2004 subscription price is just \$85 US.

---

Production and design by Angela Hanshaw, American Library Association Production Services.  
Smart Libraries Newsletter is published monthly by ALA TechSource, a unit of the publishing division of the American Library Association.  
Copyright American Library Association 2004. All rights reserved.