



Against the Grain

“Linking Publishers, Vendors and Librarians”

ISSN: 1043-2094

NISO Metasearch Initiative Targets Next Generation of Standards and Best Practices

by **Cynthia Hodgson** <chodgson@niso.org>, **Andrew Pace** (North Carolina State University) <andrew_pace@ncsu.edu>, and **Jenny Walker** (Ex Libris, Inc.) <jenny@exlibris-usa.com>

This is the first of a two-part article on the Metasearch Initiative of the National Information Standards Organization (NISO). Part 1 focuses on the issues that prompted the creation of the Metasearch Initiative and reviews NISO's plan of action. Part 2, which will appear in an upcoming issue, will review findings and recommendations of the NISO metasearch committees.

Metasearch — also called parallel search, federated search, broadcast search, and cross-database search — has become commonplace in the information community's vocabulary. All speak to a common theme of allowing search and retrieval to span multiple databases, sources, platforms, protocols, and vendors at once.

The **Z39.50** protocol has been the primary mechanism for providing metasearch since the first version of the standard was issued in 1988. This standard, which was initially designed to search across disparate library catalogs, has several drawbacks in today's metasearch environment. It was not designed for operation in a Web environment; it was not intended for article-level citations; and for many content providers it is overly complex to implement, thus creating a high barrier of entry.

Metasearch software providers have implemented a variety of protocols to ensure access to content including: **Z39.50**, IT search standards such as **SQL**, newer Web standards such as **XQuery**, customized proprietary point-to-point connections, metadata harvesting, and HTML parsing or screen scraping. This multiplicity of protocols that must be supported and the lack of commonly implemented standards, best practices, and tools make the metasearch

environment less efficient for the system provider, the content provider, and ultimately the end-user.

Metasearch Challenges

At the **2003 ALA Midwinter** meeting in Philadelphia, a group of resource providers met to discuss their concerns on the loads their systems were encountering from metasearch engines. Metasearch software agents were directing traffic toward their systems in volumes previously unseen and in a way that often caused system slowdowns. At the meeting, the **National Information Standards Organization (NISO)** offered to take a leadership role in further identifying metasearch problem areas and in proposing standards or best practice solutions.

In May 2003, **NISO** hosted a two day strategy meeting in Denver to define the metasearch-related issues and devise an action plan on ways to move forward. There was agreement that there is strong market interest in implementation of metasearching tools and that cross database search capabilities will be an area of continued growth. But metasearching has created challenges for the software providers, content providers, and implementing libraries — challenges which ultimately impact the end user. Among the issues identified were:

- **Metasearching impacts system resources and performance.**

Metasearches can spawn a large number of individual search and retrieval interactions between the meta engine and search targets, with the potential for multiple simultaneous search requests impacting a single provider's server environment.

In a Web environment, metasearching is

“stateless” meaning each search request invokes a separate authentication process. These authentication processes are resource intensive operations — vastly more intensive than mere search and retrieval operations. Additionally, problems in passing authentication information between systems and subsequent access rejections can result in users having content excluded from their metasearch, even though they have a valid license to access it.

- **Intellectual property and product branding need protection.**

Content providers have traditionally assumed that their content, whether bibliographic, citation, abstract, full text, full image, etc. would display within the provider's native interface, which conveys important information beyond the content such as “branding” and rights use declarations. Generally, such intellectual property information has not yet been embedded in individual records, so records retrieved via metasearch may not display it.

- **Competitive advantages may occur from ranking and ordering of retrieval sets.**

Content providers have concerns about how metasearch engines determine the ranking, display, and ordering of content presented to the end user. If the search engine imposes a preference or a ranking, to what degree are the content providers and the end users advantaged or disadvantaged?

Can the manner in which content is retrieved or displayed influence measure-

continued on page 80

**“Linking
Publishers, Vendors
and Librarians”**



Against the Grain

Uncommon ...

Against the Grain is your key to the latest news about libraries, publishers, book jobbers, and subscription agents. *ATG* is a unique collection of reports on the issues, literature, and people that impact the world of books, journals, and electronic information.



Against the Grain

Unconventional...

ATG is published six times a year, in February, April, June, September, November, and December/January. A six issue subscription is available for only \$40 U.S. (\$50 Canada, \$70 foreign), making it an uncommonly good buy for all that it covers. Make checks payable to *Against the Grain, LLC* and mail to:

Katina Strauch
209 Richardson Avenue
MSC 98, The Citadel
Charleston, SC 29409

Name	Address	City	State	Zip	Phone
Company	Email				

**NISO Metasearch Initiative
from page 79**

ment of use or relevance of content by either the end user or the library that purchases content services?

• Libraries need to position their services alongside free Web services.

To many end users, metasearching refers to the use of an Internet search engine such as **Google**, or a metacrawler such as **Dogpile**, which simultaneously searches multiple Internet search engines and combines the results.

Libraries offer access to numerous content services that are completely unavailable through any free Internet search engine. End users want access to this value-added information, but they have no interest in understanding the differences in how the services are offered or in learning multiple access methods.

Increasingly, libraries want to be a “portal” for their patrons into information of all kinds, both within and without the library walls, whether owned, licensed, or free to use.

In October 2003, **NISO** held an educational workshop about metasearch in Washington, D.C. where librarians, software providers, content providers, and aggregators could interact to discuss the current state of metasearch and further scope the areas that **NISO’s Metasearch Initiative** should address.



NISO Action Plan

Following the two workshops, **Jenny Walker (Ex Libris, Inc.)** and **Andrew Pace (North Carolina State University)** were asked to jointly lead the **Metasearch Initiative** with the goal of enabling:

- metasearch service providers to offer more effective and responsive services,
- content providers to deliver enhanced content and protect their intellectual property, and
- libraries to deliver services that distinguish their offerings from **Google** and other free Web services.

Three task groups were formed to pursue different aspects of the metasearch challenges.

1. Access Management

Chaired by **Michael Teets (OCLC, Inc.)**, the Access Management task group is charged with gathering requirements for metasearch authentication and access needs, inventorying existing authentication processes now in place, and developing a series of formal use cases describing the needs. The problem they want to solve is how best to certify a user from the organization authenticator to the data provider, by way of the metasearch provider, in such a

way that the authentication can be trusted end-to-end and ultimately deliver the services to which the user is entitled.

2. Collection and Service Descriptions

Chaired by **Juha Hakala (Helsinki University Library)**, the Collection and Service Descriptions task group is developing a metadata element set for collection-level description, and methods for describing informational services that are used to provide access to collections. Once the metadata element sets (semantics) and appropriate encodings (syntax) for the two areas have been specified, the Task Group will concentrate on creating a draft standard, which will serve as a basis for future rules for describing collections and services.

3. Search and Retrieval

Co-chaired by **Katherine Kott (Digital Library Federation)** and **Sara Randall (Endeavor Information Systems)**, the Search and Retrieval task group is working three areas: current metasearch practices including a standard vocabulary, citation level data elements, and metadata returned about result sets. Their committee is also developing a **Metasearch XML Gateway (MXG)** as a low-entry-barrier method for service providers to expose content to metasearch engines.

Survey of Content and System Providers

To further scope and understand the problem, the Search and Retrieval Task Group conducted a survey of content providers and library system vendors on the current state of metasearching. Key results of the survey were:

- 83 percent are aware of current metasearching activity on their database(s).
- 54 percent do not have a policy regarding metasearching of their offerings.
- Of those who do have a policy, 30 percent do not allow metasearching of their database(s).
- 54 percent believe that allowing metasearching of their offerings is very important to their customers.
- Of those who allow metasearching of their offerings, 70 percent think standards and guidelines in metasearching would be very important to their business.
- Many different search and retrieval protocols are in use, with many providers supporting more than one access method. These include: HTTP/HTML based (76%); Z39.50 (64%); XML/SOAP (33%); SQL (30%); and legacy system and/or Telnet-based access (25%).
- The most common format for display of search results was as an HTML page (84%), followed by MARC 21 (63%),

continued on page 81

SIDEBAR
NISO Metasearch Initiative Participants

Amira Aaron
Harvard University Library

Grace Agnew
Rutgers University Libraries

Katie Anstock
formerly of Talis Information Ltd.

Kristina Aston
Library and Archives Canada

Julie Blume Nye
Fretwell-Downing, Inc.

Patricia Brennan
Thomson Scientific

Mary Bushing
Library Consultant and Educator

Susan Campbell
College Center for Library Automation

Frank Cervone
Northwestern University

Robina Clayphan
The British Library

Paul Cope
Auto-Graphics, Inc.

Ray Denenberg
Library of Congress

Dana Dietz
OCLC, Inc.

Larry Dixon
Library of Congress

Matthew Dovey
University of Oxford

Emily Fayen
MuseGlobal, Inc.

Susan Farris
National Library of Medicine

Riccardo Ferranti
Smithsonian Institution Archives

David Fiander
University of Western Ontario

Liz Finlayson
MuseGlobal, Inc.

Matt Goldner
OCLC, Inc.

Cary Gordon
The Cherry Hill Co.

Reynold Guida
Thomson Scientific

Juha Hakala (TG2 Chair)
Helsinki University Library

Sebastian Hammer
Index Data

Mary Jackson
Association of Research Libraries

Pete Johnson
UKOLN, University of Bath

Anne Karle-Zenith
Michigan State University

Ted Koppel
Ex Libris, Inc.

Katherine Kott (TG3 co-chair)
Digital Library Federation

Marc Krellenstein
Elsevier

Kathy Kwan
National Library of Medicine

Ralph LeVan
OCLC, Inc.

John Little
Duke University

Eric Lochstet
Thomson Scientific

Doug Loynes
OCLC, Inc.

Mike McKenna
California Digital Library

Vicki Miller
OCLC, Inc.

Ron Miller
The H. W. Wilson Company

William Mischo
University of Illinois at Urbana-
Champaign

Alistair Morrison
Lexis Nexis Academic & Library
Solutions

Peter Murray
OhioLINK

Mark Needleman
SIRSI Corp.

Peter Noerr
MuseGlobal, Inc.

Audrey Novak
Yale University

Andrew Pace (MI co-chair)
North Carolina State University

Judith Pearce
National Library of Australia

Oliver Pesch
EBSCO Information Services

Sara Randall (TG3 co-chair)
Endeavor Information Systems

Ed Riding
DYNIX Corp.

Chris Roberts
Ex Libris, Inc.

Simona Rollinson
Follett Software Co.

Robert Sanderson
University of Liverpool

Juliane Schneider
NYU School of Medicine

Ezra Schwartz
ArtandTech.com

Ralph Scott
U. S. DOE, Office of Scientific and
Technical Information

Tim Shearer
University of North Carolina at Chapel
Hill

Sarah L. Shreeves
University of Illinois at Urbana-
Champaign

Jeff Steinman
Lexis Nexis Academic & Library
Solutions

Patricia Stevens
Consultant

Michael Teets (TG1 chair)
OCLC, Inc.

Roy Tennant
University of California

Chuck Thomas
Florida State University Libraries

Theo van Veen
National Library of the Netherlands

Jenny Walker (MI co-chair)
Ex Libris, Inc.

Terry Willan
Talis Information Ltd.

David Yakimischak
formerly of JSTOR

William Ying
ARTstor

Johan Zeeman
RLG

Candy Zemon
Polaris Library Systems

Maja Zumer
National and University Library of
Slovenia

proprietary XML (53%), Dublin Core (26%), and GRS-1 (21%). Although RSS and WSDL (Web Services Description Language) are not used by most survey respondents today, 20% indicated plans for future support.

- Respondents cited several benefits for allowing customers metasearch access: an increased customer base (79%), gaining a competitive edge (58%), and opportunities for partnership (53%).
- The main concerns of content providers with metasearch were: loss of control over search results (53%), loss of branding (53%), digital rights management (47%), customer support problems (42%), excessive use of system resources (37%), and the amount of communications required with other suppliers (21%).

The survey results were used by all three Task Groups in further refining their work plans and in developing use cases.

Next Steps

With a mix of librarians, software providers, and content providers, the three task groups have drawn the participation of over 60 individuals from five countries. (See the sidebar for the list of **Metasearch Initiative** participants and their organizations.) Each group's first set of deliverables and recommendations was presented at **NISO's** fall workshop in September 2005.

Part 2 of this article, which will appear in an upcoming issue of *Against the Grain*, will report on the **NISO Metasearch Initiative** task groups' initial set of findings and recommendations. Official documents are posted on the **NISO Metasearch Initiative** Webpage (http://www.niso.org/committees/MS_initiative.html). Committee activities can be followed at the task groups' **WIKI** (<http://www.lib.ncsu.edu/niso-mi/>). 