February 2006

NISO Metasearch Initiative Targets Next Generation of Standards and Best Practices

Cynthia Hodgson

NISO, chodgson@niso.org

Andrew Pace

North Carolina State University, andrew_pace@ncsu.edu

Jenny Walker

Ex Libris, Inc., jenny@exlibris-usa.com

Follow this and additional works at: <https://docs.lib.purdue.edu/atg>

Part of the [Library and Information Science Commons](#)

Recommended Citation

Hodgson, Cynthia; Pace, Andrew; and Walker, Jenny (2006) "NISO Metasearch Initiative Targets Next Generation of Standards and Best Practices," *Against the Grain*: Vol. 18: Iss. 1, Article 32.

DOI: <https://doi.org/10.7771/2380-176X.4861>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

The Charleston ADVISOR

Critical Reviews of Web Products for Information Professionals

"The Charleston Advisor serves up timely editorials and columns, standalone and comparative reviews, and press releases, among other features. Produced by folks with impeccable library and publishing credentials ...[t]his is a title you should consider..."

— *Magazines for Libraries*, eleventh edition, edited by Cheryl LaGuardia with consulting editors Bill Katz and Linda Sternberg Katz (Bowker, 2002).

- Over 250 reviews now available
- Web edition and database provided with all subscriptions
- Unlimited IP filtered or name/password access
- Full backfile included
- Comparative reviews of aggregators featured
- Leading opinions in every issue

\$295.00 for libraries
\$495.00 for all others

☒ **Yes! Enter My Subscription For One Year.** ☐ Yes, I am Interested in being a Reviewer.

Name _____ Title _____
Organization _____
Address _____
City/State/Zip _____
Phone _____ Fax _____
Email _____ Signature _____

Innovations Affecting Us from page 78

reserves, will they also become responsible for podcasts of course related content? At the very least, the library may find itself responsible for teaching users about podcasting and how to use the new resources that are available at their institution through podcasts. Perhaps the library will also find itself checking out digital audio players with course content to students who wish to hear a lecture but do not have a computer or Internet connection of their own. One way that academic libraries can incorporate coursecasting is to provide access to the current day or week's worth of course podcasts on the library Website. Also, libraries could further supplement coursecasting by building Web pages that not only have access to the podcasts for a particular course, but also include links to relevant library resources and possibly podcasts on using those resources. For example, on a page with access to the podcasts for a particular biology course, the library could also include links to the subject guides for biology as well as any other information on using library resources. This could, of course, become difficult to manage as more course lectures become available as podcasts.

The popularity of podcasting as a method of disseminating information has grown rapidly in the past year. As an evolving technology, podcasting (and coursecasting) is not without

continued on page 85

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 de-

signed 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.

continued on page 80

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 measurement 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.

continued on page 81

SIDEBAR
NISO Metasearch Initiative Participants

Amira Aaron Harvard University Library	Mary Jackson Association of Research Libraries	Sara Randall (TG3 co-chair) Endeavor Information Systems
Grace Agnew Rutgers University Libraries	Pete Johnson UKOLN, University of Bath	Ed Riding DYNIX Corp.
Katie Anstock formerly of Talis Information Ltd.	Anne Karle-Zenith Michigan State University	Chris Roberts Ex Libris, Inc.
Kristina Aston Library and Archives Canada	Ted Koppel Ex Libris, Inc.	Simona Rollinson Follett Software Co.
Julie Blume Nye Fretwell-Downing, Inc.	Katherine Kott (TG3 co-chair) Digital Library Federation	Robert Sanderson University of Liverpool
Patricia Brennan Thomson Scientific	Marc Krellenstein Elsevier	Juliane Schneider NYU School of Medicine
Mary Bushing Library Consultant and Educator	Kathy Kwan National Library of Medicine	Ezra Schwartz ArtandTech.com
Susan Campbell College Center for Library Automation	Ralph LeVan OCLC, Inc.	Ralph Scott U. S. DOE, Office of Scientific and Technical Information
Frank Cervone Northwestern University	John Little Duke University	Tim Shearer University of North Carolina at Chapel Hill
Robina Clayphan The British Library	Eric Lochstet Thomson Scientific	Sarah L. Shreeves University of Illinois at Urbana- Champaign
Paul Cope Auto-Graphics, Inc.	Doug Loynes OCLC, Inc.	Jeff Steinman Lexis Nexis Academic & Library Solutions
Ray Denenberg Library of Congress	Mike McKenna California Digital Library	Patricia Stevens Consultant
Dana Dietz OCLC, Inc.	Vicki Miller OCLC, Inc.	Michael Teets (TG1 chair) OCLC, Inc.
Larry Dixon Library of Congress	Ron Miller The H. W. Wilson Company	Roy Tennant University of California
Matthew Dovey University of Oxford	William Mischo University of Illinois at Urbana- Champaign	Chuck Thomas Florida State University Libraries
Emily Fayen MuseGlobal, Inc.	Alistair Morrison Lexis Nexis Academic & Library Solutions	Theo van Veen National Library of the Netherlands
Susan Farris National Library of Medicine	Peter Murray OhioLINK	Jenny Walker (MI co-chair) Ex Libris, Inc.
Riccardo Ferranti Smithsonian Institution Archives	Mark Needleman SIRSI Corp.	Terry Willan Talis Information Ltd.
David Fiander University of Western Ontario	Peter Noerr MuseGlobal, Inc.	David Yakimischak formerly of JSTOR
Liz Finlayson MuseGlobal, Inc.	Audrey Novak Yale University	William Ying ARTstor
Matt Goldner OCLC, Inc.	Andrew Pace (MI co-chair) North Carolina State University	Johan Zeeman RLG
Cary Gordon The Cherry Hill Co.	Judith Pearce National Library of Australia	Candy Zemon Polaris Library Systems
Reynold Guida Thomson Scientific	Oliver Pesch EBSCO Information Services	Maja Zumer National and University Library of Slovenia
Juha Hakala (TG2 Chair) Helsinki University Library		
Sebastian Hammer Index Data		

NISO Metasearch Initiative from page 81

- 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%), 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/>).

Technology Left Behind — What in the WorldCat is OCLC up to?

Column Editor: **Cris Ferguson** (Electronic Resources/Serials Librarian, James B. Duke Library, Furman University, 3300 Poinsett Highway, Greenville, SC 29613; Phone: 864-294-2713) <cris.ferguson@furman.edu>

In October 2003, OCLC launched the pilot of the **Open WorldCat** program, an initiative which “makes records of library-owned materials in OCLC’s WorldCat database available to Web users on popular Internet search, bibliographic and bookselling sites.” (<http://www.oclc.org/worldcat/open/default.htm>) The program has been expanding ever since, leaving the pilot phase in October 2004.

Since making **Open WorldCat** a permanent program at the end of 2004, OCLC has continued to develop and to enhance the functionality of the program. On January 3, 2006, OCLC announced the acquisition of the assets of **Openly Informatics**, opening the door for even further expansion of services. This column will investigate how **Open WorldCat** works, recent developments in the program, and future directions for growth, touching on the recent acquisition of **Openly Informatics**.

Open WorldCat Partnerships

The goal of the **Open WorldCat** program is to make OCLC member libraries more visible on the Web and their collections more accessible to Web users. As part of this effort, OCLC has established linking partnerships with several vendors, including search engines, book vendors, and fulfillment services. Through the partnerships, **Open WorldCat** allows these vendors access to “millions of abridged WorldCat records.”

Both **Google** and **Yahoo! Search** are partnering with OCLC to deliver links to **WorldCat** records. According to **Chip Nilges**, Vice President, OCLC New Product Planning,

“**Yahoo! Search** and **Google** are exposing a 3.4 million record set through their interfaces, including the 3 million most widely owned items in **WorldCat**,” representing 78 percent of the holdings in the **WorldCat** database.

How Open WorldCat Works

To find items held by a local library, a user conducting a search in **Google** or **Yahoo! Search** enters a search phrase that matches the title of the desired item and the phrase “find in a library.” At this time, the “find in a library” search online searches the title field of a record. It will not search for an author or subject. When the results of a search are displayed, links to **Open WorldCat** records are marked by the phrase “Find in a Library.”

Clicking on the “Find in a Library” link in the search results takes the user into the **Open WorldCat** record for the desired item. Once in the **Open WorldCat** record, the user has the option to enter his or her zip code to determine if a local library owns the item. If the book is available at a local library, the record provides a link to the library’s hours, if available, and also a link to the online catalog, letting the user check to see if the book is checked in and available.

An example of how this entire process looks to the user can be found on the **Open WorldCat** Web page <http://www.oclc.org/worldcat/open/how/default.htm>.

While in the pilot phase, **Open WorldCat** was open to participation by all libraries with holdings in **WorldCat**. As an official program, participation has been limited to those libraries that subscribe to **WorldCat** on the **FirstSearch** platform. As **Nilges** points out, “The program requires resources to keep running (and improving). We have added it as one of the many features an institution receives as a subscriber to **WorldCat** on **FirstSearch**, and, by participating, institutions help us put more resources into the project to encourage further innovations and features.” Those libraries that do not subscribe to **WorldCat** on **FirstSearch** will not be able to open their holdings up to users via **Open WorldCat**.

Toolbars

The primary disadvantage to using the “find in a library” method to search for and to access **Open WorldCat** records is that the user has to know to add the “find in a library” phrase to the search string. To make accessing **Open WorldCat** a little more user friendly, OCLC has worked with its partners to create Web-based search tools, including a special OCLC edition of the **Yahoo! Search** toolbar. The **Yahoo! Search** toolbar, while still equipped with other popular features like **Anti-Spy** and **Pop Up Blocker**, also includes a feature that allows users to search for terms in **Open WorldCat**. Instructions for downloading and using the

continued on page 83