

NISO Metasearch Initiative

Content Provider Survey

Standards Committee BC / Task Group 3

October 2004



Contents

Introduction	1
Survey Results	1
Survey Questions.....	2
Glossary.....	6

Metasearch Content Provider Survey

Introduction

To scope and understand the metasearch problem, the NISO Metasearch Initiative, Search and Retrieval Task Group, conducted a survey of content providers and library system vendors on the current state of metasearching. The survey was completed in October 2004.

This document contains a summary of the survey results, followed by a copy of the survey questionnaire.

In the process of creating the survey questionnaire the task group developed a glossary of related terms. This glossary is reproduced following the survey questionnaire.

Survey Results

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. HTTP/HTML based (76%); Z39.50 (64%); XML/SOAP (33%); SQL (30%); legacy system and/or Telnet-based access (25%).
- The most common format for display 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%).

Metasearch Content Provider Survey

Survey Questions

SURVEY INTRODUCTION:

Thank you for taking the time to complete this important survey. Your response will provide the factual foundation that will allow us to develop metasearch standards and practices.

Questions:

1. How familiar were you with the NISO Metasearch Initiative before receiving the email about this survey?
 - I had not heard of it before
 - I had heard of it but did not know much about it
 - I was already following this initiative
2. Are you aware of current metasearching activity on your database(s)?
 - Yes
 - No
3. Does your organization have a policy regarding metasearching of your offerings?
 - Yes
 - No → SKIP TO Q6
4. Does your policy allow metasearching against your database(s)?
 - Yes, metasearching is allowed → SKIP TO Q6
 - No, metasearching is not allowed → ASK Q5 ONLY IF Q4 = "NO"
5. Do you plan to support metasearching in any form within the next year?
 - Yes
 - No → SKIP TO Q7
6. Does your policy allow or require:
 - A stateful session
 - Timeout
 - Additional fees
 - Other (specify _____)
 - None of the above
7. How important do you believe allowing metasearching of your offerings is to your customers?
 - Very important
 - Somewhat important
 - Not very important
 - Not at all important
 - Not applicable
8. Have any of your customers requested that the database(s) offered by your organization be made available through metasearch technology?
 - Yes, we have received many requests like this
 - Yes, we have received some requests like this
 - No, we have not received requests like this

If Q7 is NA and Q8 is NO, skip to END

Metasearch Content Provider Survey

9. Which of the following search and retrieval methods do you currently support, or plan to support within the next year?

	Currently Support	Plan to support	Do not support/ no plans to do so	Don't know
Z39.50				
http/html based				
XML/SOAP based				
Legacy (e.g. telnet) based				
SQL based				
Xquery/Xpath based				
Other (specify _____)				

10. Which metasearch results formats are provided by your service? (Check all that apply.) MARC21

- GRS-1
- Dublin Core
- Proprietary XML
- HTML Page
- RSS
- UDDI
- WSDL
- Other schema (e.g., MODS or a DTD). Please specify _____

11. Approximately how many person months/years were expended to create the current search and retrieval method, including specification, coding, testing, and documentation?

- Three person/months
- Six person/months
- One person/year
- Two person/years
- More than two person/years
- Don't know

12. Are you planning changes or enhancements to your current services?

- Yes/
- No. If yes, go to 13. If no, go to 15.

13. Please indicate the metasearch results formats you plan for your service. Include both enhancements to existing services and new services. (Check all that apply.)

- GRS-1
- Dublin Core
- Proprietary XML
- HTML Page
- RSS
- UDDI
- WSDL
- Other schema (e.g., MODS or a DTD). Please specify _____

14. What is the estimated number of person months/years your organization will devote to implementing or enhancing your system, including specification, coding, testing, and documentation?

- Three person/months
- Six person/months

Metasearch Content Provider Survey

- One person/year
- Two person/years
- More than two person/years
- Don't know

15. What advantage is your company gaining by allowing metasearch access for your customers? (check all that apply)

- Increased customer base
- Opportunities for partnerships
- Competitive edge
- Other (specify _____)

16. What business or technical issues is metasearch access requiring your organization to address? (check all that apply)

- Loss of branding
- Digital rights management
- Loss of control of search results
- Too much communication required with other vendors
- Support problems with customers
- Excessive use of system resources
- Other (specify _____)

17. How important do you think standards and guidelines in this area would be to your business?

- Very important
- Somewhat important
- Not very important
- Not at all important

18. In which of the following technical areas would either standards or guidelines help you? (check all that apply)

- Access management
- Metasearch identification
- Searching options
- Resource description
- Results set management
- Statistics
- Other (specify _____)

19. If new standards or guidelines are created, what level of effort are you willing to expend to implement them?

- Three person/months
- Six person/months
- One person/year
- Two person/years
- More than two person/years
- Don't know

20. For the following authentication methods, please indicate the likelihood that you will implement the method for your service in the time period listed:

Technology	In use	6 mo.	1 yr.	18 mo.	2 yrs.	Never
IP range						
Usrnm/Pwd						

Metasearch Content Provider Survey

Referrer URL						
NCIP						
Shibboleth						
Athens						
PassPort						
Liberty Alliance						
Other (specify)						

21. Please comment generally on your plans for authentication and authorization services especially as they relate to Metasearch search and retrieval.

22. In which type of organization do you work?

- Non Profit (including societies)
- Government
- Aggregator
- Primary Commercial Publisher
- Academic Press
- Library Management System Vendor
- Other (specify _____)

23. Please complete the following information about you (optional):

Name _____
Title _____
Email address _____
Phone number _____

Thank you for responding to this survey!

Metasearch Content Provider Survey

Glossary

28 May 2004

This glossary is for use with the NISO Metasearch Initiative Content Provider Survey.

Metasearch	Metasearch systems are specialized search engines that allow a user to specify a single search and have it broadcast to a large number of search engine servers. A metasearch may or may not include a variety of syntaxes based on the target search engines.
Athens	A national authentication service employed by the United Kingdom higher education system using geographically distributed and replicated servers using Athens software and the Athens national user information database. Provides single signon access to licensed resources from vendors including EBSCO, OVID, Lexis-Nexis, and ISI.
DTD	Document Type Definition for an XML document. May reside within an XML document or external to it. Defines the document structure and its allowable or required components of an XML file.
Dublin Core	Refers to Dublin Core Metadata Element Set produced by the Dublin Core Metadata Initiative. XML-based standard began as a core set of descriptive elements that could be used across a broad range of disciplines and vertical industries. http://dublincore.org/
GRS	Generic Record Syntax also known as GRS-1. A structured record syntax specified in the Z39.50 standard as an acceptable record syntax for retrieval records representing a database record. Contains a logical tree that describes the hierarchical structure of the abstract database record. http://www.loc.gov/z3950/agency/markup/23.html
HTML	Hypertext Markup Language . A subset of the Structured Generic Markup Language, this coded format language controls how hypertext documents appear in HTML capable applications such as web browsers and, more recently, email programs.
HTTP	Hypertext Transfer Protocol . Used to move hypertext files across the Internet. It requires HTTP client and server components, typically a web browser (client) and a web server (server). A variation, https, is used for secure transmissions.
IP	Internet Protocol . Often used as short hand for IP Address. IP is a basic protocol used to communicate on the Internet. Each address is numeric consisting of 4 three-character numbers separated by periods. Addresses are used to route Internet traffic and establish connections.

Metasearch Content Provider Survey

- Liberty Alliance** A project which has produced Web Services definitions in order to establish an open standard for federated network identity. Federated identity allows users to link elements of their personal information from distributed locations. http://www.service-architecture.com/web-services/articles/liberty_alliance_project.html
- MARC** **Machine Readable Cataloging**, or Z39.2. Multiple variations of this standard exist, all of which serve to represent and communication bibliographic and related information in machine readable form. The dominant form is MARC21 used in the United States, Canada, and Australia. Other notable variations include UniMARC (Western Europe), and CNMARC (China) as well as Scandinavian implementations. <http://www.loc.gov/marc> and <http://www.niso.org/standards/resources/z39-2.pdf>
- MODS** **Metadata Object Description Schema**. This XML-based schema establishes a protocol for encoding bibliographic information using natural language identifiers. Intended to represent selected information from MARC records. May be used in isolation or embedded in METS documents. <http://www.loc.gov/standards/mods> .
- NCIP** **National Circulation Interchange Protocol**, or Z39.83-2002. This protocol exchanges messages between applications for the purpose of lending and borrowing items, and providing controlled access to electronic resources. <http://www.niso.org/standards/resources/z3983pt1rev1.pdf>
- Passport** The Microsoft .NET centralized authentication service offers a single logon and core profile services for member websites. <http://www.passport.net/Consumer/Default.asp?lc=1033>
- RSS** **Really Simple Syndication**, a group of XML-based communication protocols that facilitate web syndication. Most frequently used by news websites and weblogs. Allows web developers to make a portion of their website content available to be repackaged into other computer programs. Example: news feed email program that repackages new articles and article updates into emails including link directly to the website content. Primary protocols are: *RDF Site Summary* and *Rich Site Summary*.
- Schema (XML)** The expression of shared vocabularies which provide a means for defining structure, content and semantics of XML documents. XML schema are quickly becoming more popular than DTDs because of the additional flexibility they provide. <http://www.w3.org/XML/Schema>
- Shibboleth** Developed as a project of the Internet 2/MACE, it is an authentication system and architecture that facilitates resource sharing between institutions. <http://shibboleth.internet2.edu/>
- SOAP** **Simple Object Access Protocol**. A protocol for XML-based remote procedure calls (RPC) over HTTP. Allows direct communication between applications over the Internet without inadvertent blocking by firewalls. <http://www.w3.org/TR/soap/>

Metasearch Content Provider Survey

SUTRS	Simple U nstructured T ext R ecord S yntax, specified as an acceptable record syntax in Z39.50. Used in a Search or Present response. No elements are explicitly defined in a SUTRS record; it is a string of textual data terminated by ASCII LF (X'0A'). http://lcweb.loc.gov/z3950/agency/Z39-50-2003.pdf
UDDI	U niversal D escription, D iscovery and I ntegration. A specification for standardized directories of information about web services including requirements, location, and capabilities.
URL	U niversal R esource L ocator is the address of a site on the internet. Often begins http:// but can also be used for FTP, Gopher, Telnet, etc...
WSDL	W eb S ervices D escription L anguage. This XML format describes network services as sets of endpoints used in document-focused or procedure-focused messages. Used in relation to UDDI, SOAP, HTTP GET/POST, and MIME to represent the capabilities of the specific XML Web Service.
XML	e X tensible M arkup L anguage is a flexible text format derived from SGML to meet the challenges of large-scale electronic publishing. It has since become a very popular format for material both on the Web and in back office environments. http://www.w3.org/XML/
Xpath	XML Path Language used for addressing parts of an XML document to be used by XSLT and XPointer. Largely consists of location expressions and paths. http://www.w3.org/TR/xpath
Xquery	Used in conjunction with Xpath, this is an XML query syntax similar to SQL designed to be broadly applicable. http://www.w3.org/TR/xquery/
Z39.50	A national standard that is popular internationally as well, Z39.50 defines communication for computer-to-computer information retrieval. Enables searchers to query other systems without knowledge of the remote systems' native searching syntax. http://www.loc.gov/z3950/agency/