

NISO RP-2006-02

NISO Metasearch Initiative

Metasearch XML Gateway Implementers Guide Version 1.0

A Recommended Practice of the National Information Standards Organization

Standards Committee BC / Task Group 3

August 7, 2006

**Published by the National Information Standards Organization
Bethesda, MD**



Contents

Foreword	ii
1 Purpose and Audience.....	1
2 Overview	1
3 Levels of Implementation	2
4 Prerequisites	3
5 Decision Points	3
6 The MXG Protocol	4
6.1 MXG URL Request	4
6.1.1 Syntax	4
6.1.2 MXG Request Parameters.....	4
6.1.3 Parsed Examples	5
6.1.4 Result Set IDs	6
6.2 MXG XML Response	6
6.2.1 MXG Response Parameters	7
7 Compliance	12
7.1 URL Request Compliance.....	12
7.2 MXG Response Compliance.....	13
8 Advanced Interoperability (Levels 2 and 3)	13
8.1 Explain (Level 2 Compliance)	13
8.2 CQL (Level 3 Compliance).....	14
Appendix A : Implementation Help.....	15
Appendix B : Resources.....	16
Appendix C : Glossary	18

Tables

Table 1: MXG request parameters.....	5
Table 2: MXG response header subelements	7
Table 3: MXG response result set elements.....	8
Table 4: MXG response record elements	9
Table 5: MXG response browser elements.....	9
Table 6: MXG response diagnostic elements	10

Figures

Figure 1: High level model of metasearch environment	1
Figure 2: MXG protocol model	2

Foreword

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 and integrate the results. Metasearch services rely on a variety of approaches to search and retrieval including open standards (such as NISO's Z39.50), proprietary API's, and screen scraping (extracting information from HTML responses). However, the absence of widely supported standards, best practices, and tools makes the metasearch environment less efficient for the metasearch service provider, the content provider, and ultimately the end-user.

NISO Metasearch Initiative

To move toward industry solutions, NISO sponsored a Metasearch Initiative to enable:

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

The groundwork for NISO's Metasearch Initiative was laid in two important events:

- A two day strategy meeting in May 2003 defined the metasearch state-of-the-art and built consensus on ways to move forward.
- A Metasearch workshop in October 2003 informed librarians, content providers, and aggregators about metasearch.

Following these meetings, NISO established three Task Groups / Standards Committees to address the different Metasearch needs areas:

- **Access Management** (Standards Committee BA / Task Group 1)
- **Collection and Service Descriptions** (Standards Committee BB / Task Group 2)
- **Search and Retrieval** (Standards Committee BC / Task Group 3)

Search and Retrieval Task Group

The Search and Retrieval Task Group was charged with identifying and/or developing standards and best practices to improve interoperability between metasearch services and content providers. In particular, they were asked to investigate and report on:

- Current practices in metasearching search and retrieval
- Common metasearch vocabulary and terms
- Result set data elements
- XML interfaces
- Best practices for metasearch search and retrieval

This document represents one of the deliverables of the Search and Retrieve Task Group. It describes how to implement a Metasearch XML Gateway (MXG) that will allow a content provider's resource to be accessed by a Metasearch Service and included in the results of a metasearch.

Background on SRU

In the late 1990s, there was significant interest in making the Z39.50 standard more relevant to the Web environment. The result was the development of a new specification: SRU (Search / Retrieve via URL)

Metasearch XML Gateway Implementers Guide

The specification retains such Z39.50 concepts as result sets, abstract record schemas, application level diagnostics, and "Explain", but differs from Z39.50 in the use of Web interfaces, XML, and the Common Query Language (CQL).

The Metasearch Initiative Search / Retrieve Task Group selected SRU as the foundation on which to build the NISO Metasearch XML Gateway (MXG). For further information about SRU and CQL, see the [Resources Appendix](#).

Metasearch Search and Retrieve Task Group

Members of the Task Group included:

Katherine Kott, co-chair
Digital Library Federation

Sara Randall, co-chair
Endeavor Information
Systems

Amira Aaron
Harvard University Library

Paul Cope
Auto-Graphics, Inc.

Ray Denenberg
Library of Congress

Dana Dietz
OCLC, Inc.

Matthew Dovey
University of Oxford

Susan Fariss
National Library of
Medicine

Riccardo Ferrante
Smithsonian Institution
Archives

Matthew Goldner
OCLC, Inc.

Cary Gordon
The Cherry Hill Co.

Renny Guida
Thomson Scientific

Sebastian Hammer
Index Data

Mary Jackson
Association of Research
Libraries

Marc Krellenstein
Elsevier

Ralph LeVan
OCLC, Inc.

John Little
Duke University

Mike McKenna
California Digital Library

Ron Miller
The H.W. Wilson
Company

William Mischo
University of Illinois,
Urbana-Champaign

Peter Murray
OhioLink

Peter Noerr
MuseGlobal, Inc.

Audrey Novak
Yale University

Andrew Pace
North Carolina University

Oliver Pesch
EBSCO Information
Services

Chris Roberts
Ex Libris, Inc.

Simona Rollinson
Follett Software Co.

Robert Sanderson
University of Liverpool

Ezra Schwartz
ArtandTech.com

Jeff Steinman
Lexis Nexis Academic &
Library Solutions

Patricia Stevens, NISO
SDC Liaison
Consultant

Roy Tennant
University of California

Jenny Walker
Ex Libris, Inc.

David Yakimischak
formerly with JSTOR

Johan Zeeman
Research Libraries Group

Candy Zemon
Polaris Library Systems

Metasearch XML Gateway Implementers Guide

Metasearch XML Gateway Implementers Guide

1 Purpose and Audience

The NISO Metasearch XML Gateway (MXG) is a low-barrier-to-entry method to expose content to metasearch services and more effectively interoperate with metasearching applications. The MXG protocol defines a simple message and response which allows a metasearch service to query a content database and receive a standardized XML response.

The MXG protocol is primarily directed to Content Providers who wish to expose their resources to one or more metasearch providers without expending substantial development resources. Metasearch Providers who will be using a Content Provider's MXG to access resources will also find this guide useful.

2 Overview

The NISO Metasearch XML Gateway (MXG) provides a mechanism for you, as a content provider, to expose your content and services to a Metasearch Service. Metasearch Services are a class of services that allow an end user to find content in multiple services with a single search. (For background information on metasearch, see the [Resources Appendix](#).) Figure 1 provides a high-level model of the Metasearch environment.

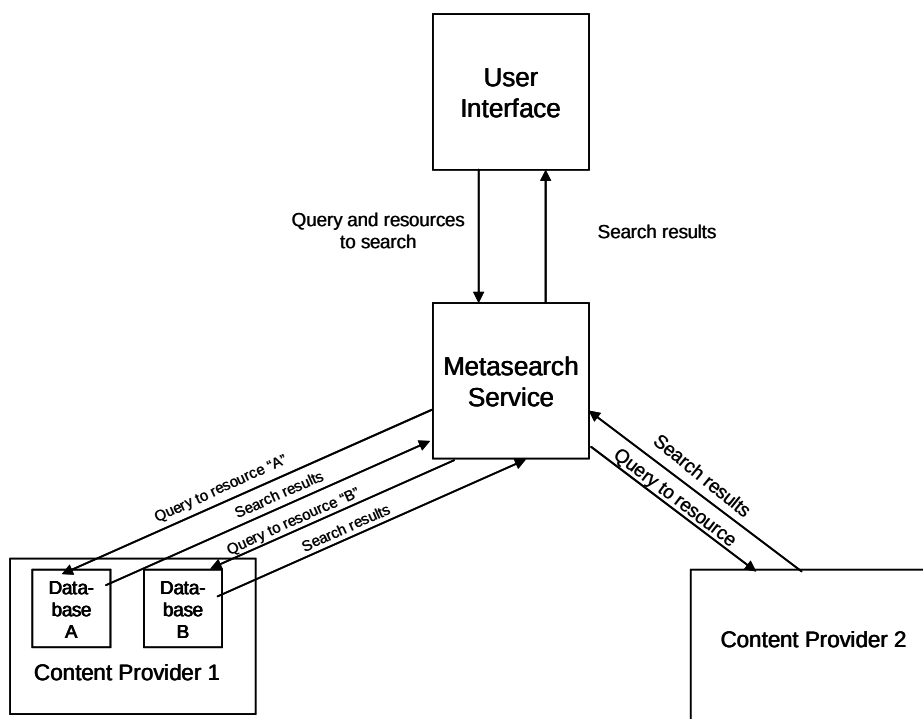


Figure 1: High level model of metasearch environment

The NISO Metasearch XML Gateway (MXG) protocol is a mechanism that can provide the query and the search results between the Metasearch Service and the resource as represented in this high level model. MXG a simple message / response based on the NISO-registered Search and Retrieve URL (SRU)¹ protocol. The Metasearch Provider sends individual queries for each resource that you host using MXG

¹ SRU utilizes the Z39.50 semantics to provide search and retrieval in a Web environment. See the [Resources Appendix](#) for more information on SRU

URLs via HTTP (Hypertext Transport Protocol). You, the Content Provider, return an MXG compliant XML formatted response to those queries. The Metasearch Service is responsible for parsing, aggregating, and displaying of the records retrieved from multiple sources—which may include records retrieved by methods other than the MXG gateway—to the end user.

The MXG protocol utilizes a standard client/server model where the Metasearch Service acts as the "client" and your database is the "server". (For basic information about the client/server model, see the [Resources Appendix](#).) Figure 2 illustrates the MXG protocol.

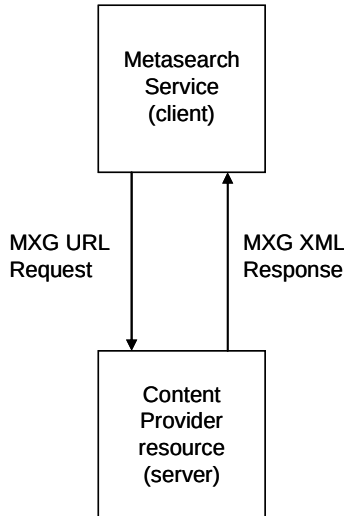


Figure 2: MXG protocol model

The steps involved in an MXG transaction are:

1. The Metasearch Service (MS) transmits an MXG URL request that incorporates the user's search query to a Content Provider's (CP) resource, usually a database.
2. The CP server receives and interprets the MXG URL and conducts the specified search query on its database.
3. The search query results are packaged into an MXG XML response, which is transmitted back to the MS.

3 Levels of Implementation

Three levels of implementation are defined for MXG. Each level requires increasing compliance with specifications of the SRU protocol; only the third level is fully compliant SRU. Full (Level 3) compliance is **strongly** encouraged for all implementers as interoperability and functionality increase with each level of implementation.

- **Level 1 defines a standard URL which will accommodate any query syntax or language.**
With Level 1 compliance, the Metasearch Service will have to convert its users' queries to the Content Provider's native search language. The amount of customization required depends on how proprietary the search language is.
- **Level 2 extends Level 1 by adding the requirement that servers provide an SRU Explain record to define the capabilities of the server.**
With Level 2 compliance, the Content Provider's server would provide an XML-formatted record that includes information about the resource, such as its host name and port and the database name, which is used as the context part of a URL.

Example:

The base URL for an MXG search against this resource:

```
host=oclc.org  
port=80  
database=search/WorldCat
```

would be:

```
http://oclc.org/search/WorldCat
```

Optionally, a Level 2 compliant server also includes human-readable information about the database(s) on the server such as its name, a database description, information about the indexes available for searching the database, and the schemas that can be used to display returned records. This server / database description utilizes the SRU Explain operation. See the [Resources Appendix](#) for background information on SRU Explain.

Section 8.1 in Advanced Interoperability further defines the use of the Explain function in Level 2 of MXG.

- **Level 3 extends Level 2 by adding the requirement that servers support a standard query grammar: CQL (Common Query Language).**
"CQL tries to combine simplicity and intuitiveness of expression for simple, every day queries, with the richness of more expressive languages to accommodate complex concepts when necessary."²
Support for CQL eliminates the need for any customized search interface, which could make the content more widely available to services that don't have the resources to write custom interfaces. See the [Resources Appendix](#) for background information on CQL.

Section 8.2 in Advanced Interoperability further defines the use of CQL in Level 3 of the MXG.

4 Prerequisites

There are very few prerequisites for implementing MXG. Obviously you will need to have an electronic content resource that is Internet accessible. This resource must have:

- a Web-addressable Uniform Resource Identifier (URI),
- a server that can receive and parse a URL request into its components parts: the base URL and parameter names and values,
- an existing search interface, and
- the ability to output search results in XML.

If you do not currently output search results in XML format but do have some type of API to a web interface for displaying records, then implementing the XML MXG response should be fairly easy. For further information on XML, see the [Resources Appendix](#).

5 Decision Points

Prior to implementing the MXG, you will need to make two decisions:

1. What level of compliance will you implement?

See the Levels of Implementation section above for the description of each compliance level. While Level 1 requires the least amount of standardized protocol use and is the easiest for the Content Provider to implement, it requires the Metasearch Provider to have or create a custom interface to your search language. Any changes to your search language could necessitate coordinated changes by the Metasearch Provider to ensure that your content continues to be retrieved accurately. The

² Common Query Language, CQL Version 1.1 13th February 2004. <<http://www.loc.gov/standards/sru/cql/index.html>>