Institute of Museum and Library Services National Leadership Grant

---

Realizing the Vision of

Networked Access to Library Resources

An Applied Research and Demonstration Project to Establish and Operate a Z39.50 Interoperability Testbed

Phase 2 – A Radioactive MARC Record Approach to Interoperability Testing



Interim Report to The Institute of Museum and Library Services

January 1 through June 30, 2003

William E. Moen, Ph.D.
<wemoen@unt.edu>
Principal Investigator
School of Library and Information Sciences
Texas Center for Digital Knowledge
University of North Texas
Denton, TX 76203

July 1, 2004

Introduction

This document provides an interim report on the Z39.50 Interoperability Testbed Project (Z-Interop) covering the period of January 1, 2003 through June 30, 2003. In February 2004, the Principal Investigator (PI) submitted to IMLS a request for extension on this project. A notice of approval from IMLS arrived in June. This document highlights activities and accomplishments to communicate to IMLS progress on our project since the interim report on January 1, 2004. Given the date of approval for the extension, this report primarily reflects activities during the latter half of June 2004.

Accomplishments and Challenges

This section summarizes key accomplishments. Subsequent sections discuss these in more detail.

Accomplishments

- Submitted and received approval for an extension to the project
- Engaged in startup activities for Phase 2 of the project
- Drafted *Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records*
- Drafted *Project Work Plan*
- Identified and secured project personnel.

Project Personnel

The Phase 2 project team consists of:

- The PI
- A masters student working 20 hours a week
- A Ph.D. student working 10 hours a week.

In addition, the project is subcontracting with a consultant and a software firm:

- Penelope Benardino, a recent graduate from the of the School of Library and Information Sciences who worked on the project in 2003
- Index Data, a software firm specializing in Z39.50 that was listed in the request for extension to provide technical support for the project.

The PI has been working to get personnel started on project work as of July 1.

Project Management

The PI has overall project management responsibilities for scheduling work, keeping the project team members on task, disseminating information about the project, and

overseeing the design and progress of the project. The PI communicates regularly with the project team via email and meetings.

The PI drafted a ***Project Work Plan*** (see Attachment A) that has been distributed to the team. The work plan identifies the major areas of activities, preliminary tasking, timeline, and deliverables. Discussions among the project team will result in modifications to the work plan. This plan will guide the work in the coming months.

The PI also drafted a document to describe the conceptual framework for the project (see Attachment B). The conceptual framework identifies the primary components of the alternative approach for interoperability testing using radioactive MARC records. As work proceeds, this framework will be enhanced and detailed.

Project Website and Information Dissemination

The project website <http://www.unt.edu/zinterop/> will serve as a primary vehicle for promoting and publicizing the project. The site will be modified and updated to reflect Phase 2 of the Z-Interop Project. The PI plans to submit announcements of the Phase 2 work to alert the community of this work.

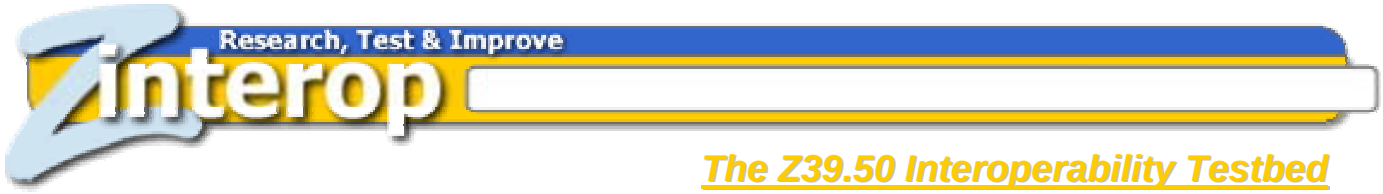
Project Technology and Software

During this start up period, the PI has been preparing the necessary computing resources for the project team.

Summary and Next Steps

Work on Z-Interop Phase 2 began with IMLS approval of the request for extension. Most of the activities have been focused on project startup. The draft Project Work Plan is under discussion by the project team and has assisted in getting the team members oriented to the project. Team members will begin as of July 1 with tasks that have been assigned. We anticipate major progress in the next three to four months in preparing the testbed for operational interoperability testing later in Fall 2004.

Attachment A: Draft Project Work Plan



The Z39.50 Interoperability Testbed

U.S. Federal Institute of Museum and Library Services
National Leadership Grant, Phase 2

A Radioactive MARC Record Approach for Interoperability Testing

---

Project Work Plan

DRAFT

William E. Moen, Ph.D.
<wemoen@unt.edu>
Principal Investigator

School of Library and Information Sciences
Texas Center for Digital Knowledge
University of North Texas
Denton, TX 76203

June 2004

Table of Contents

1. Introduction 1
2. Project Work Areas 1
   2.1 Creating MARC documentation database 1
   2.2 Creating MARC Records 2
   2.3 Developing Test Searches 2
   2.4 Creating Automated Testing Scripts 3
   2.5 Indexing Guidelines 3
   2.6 Validating Testing Procedures 4
   2.7 Analysis and Reporting 4
   2.8 Conduct Interoperability Testing with Libraries 4
3. Project Timeline 5

Project Work Plan

1. Introduction

This document details a work plan to guide the planning and execution of a new phase of the Z39.50 Interoperability Testbed Project. In this phase, we will use a set of special, diagnostic MARC records to identify interoperability problems between a Z39.50 client and a Z39.50 server providing access to a database of bibliographic records supporting the search and retrieval functions of an online library catalog. We refer to these special, diagnostic records as radioactive MARC records. The project includes a number of separate but related activities. The structure of the work plan groups these activities into work areas, each of which describes the major activities, tasks, deliverables, timelines, and responsibilities. The work areas are:

- Creating MARC documentation database
- Developing Radioactive MARC records
- Identifying test searches
- Enhancing indexing guidelines for searching
- Creating testing procedures and processes
- Validating radioactive MARC record approach for interoperability testing
- Conducting interoperability testing

A number of these work areas will proceed in parallel, and there are interactions between the work areas.

2. Project Work Areas

2.1 Creating MARC documentation database

This work area focuses on creating a MySQL database to store MARC content designation information. The purpose of this database is to be able to quickly find all content designation structures that can be used for various formats of materials. The final implementation of the database will run under Linux; it will be web accessible via PhP scripts for administration as well as data input.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Design and Prototype DB	• Identify requirements • Develop initial design • Prototype design in MySQL • Prototype user interfaces • Determine backup procedures • Develop XML export functionality	Draft database design; DB prototype	June 1-30	Jason Bill
Implement DB	• Finalize design and specifications • Implement DB under Windows for testing • Implement DB under Linux • Implement user interfaces and connect via PhP to database • Test DB application and interfaces	DB application with user interfaces for data input	June 15-July 15	Jason
Populate DB	• Identify sources of data that can be batch loaded • Identify sources of data that will be manually input • Load data		July 1-30	Jason Penny

Activities	Tasks	Deliverables	Timeline	Responsibilities
	• Input data			
Validate DB contents	• Test database with appropriate queries • Verify accuracy of data in database		July 15-30	Jason Penny
Documentation	• Write complete documentation for database application	Complete document describing DB, procedures, scripts, data loading, etc.	July 30	Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
MARC DB								

2.2 Creating MARC Records

This work area covers all the activities that are required to create the radioactive MARC records. This includes the identification of the character strings that will serve as data values for each field/subfield in a record, identifying tools for creating the records, and the actual creation and validation of the records.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Requirements Setting	• Identify requirements for data for records • Determine options for data value character strings • Account for various format specific content designation • Determine process for generating character strings • Identify number and type of records required	Draft requirements document for discussion with Index Data	June 15-July 15	Bill Penny Index Data
Create character string patterns	• Develop tool for producing character strings • Produce character strings	Set of character strings available for data input	July 15-30	Index Data
Identify MARC creation tool(s)	• Identify functionality needed to create radioactive records • Identify potential tools for record creation • Install or implement tool		July 1-15	Penny Index Data
Input data into MARC records	• Create MARC records	Set of Radioactive MARC Records	Aug 1-Sept 1	Penny Jason
Validate MARC records	• Identify procedures to validate records		Aug 1-Sept 1	Penny Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
MARC Records								

2.3 Developing Test Searches

This work area covers the activities and tasks related to identifying and agreeing on the test searches to be used. The results of this work area inform MARC records needed and addressed in 2.2.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Identify Searching Requirements	• Use U.S. National Profile for Library Applications,		July 1-15	Penny Bill

Activities	Tasks	Deliverables	Timeline	Responsibilities
	Levels 0 and 1 • Determine character string patterns needed in the record for the searches • Determine indexing requirements (see 2.5)			
Document Test Searches	• Prepare document that details each test search, search term, and expected search results	Test Searches document	July 15- Aug 30	Penny Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Test Searches								

2.4 Creating Automated Testing Scripts

This work area covers the activities and tasks related to developing and creating automated mechanisms for issuing test searches and compiling results. Index Data has primary responsibility for this work area.

Activities	Tasks	Deliverables	Timeline	Responsibilities
	•			
	•			
	•			
	•			
	•			

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Testing Scripts								

2.5 Indexing Guidelines

This work area addresses the expansion of the existing indexing guidelines developed for the first phase of interoperability testing to address the interoperability testing using the radioactive MARC records.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Create draft indexing policies	• Identify appropriate content designation to be indexed for each test search • Develop draft indexing policies • Distribute for review and comment	Draft indexing guidelines	Aug 1 – Sept 15	Penny Jason
Revise and finalize indexing policies	• Compile comments from review of draft • Clarify any questions on content designation to be indexed • Create final version of indexing policies	Indexing Guidelines for Interoperability Testing	Sept 15-30	Penny Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Indexing Policies								

2.6 Validating Testing Procedures

This work area uses the existing Z-Interop reference implementation server and information retrieval system (Sirsi Unicorn) to test and validate all interoperability testing procedures using radioactive MARC Records.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Implement indexing guidelines	• Review procedures for setting indexing policies in Sirsi system • Set up enhanced indexing policies on the system		Sept 1-15	Jason JungWon
Load records	• Use record loading procedures to load radioactive MARC records into system		Sept 15-30	Jason JungWon
Test automated scripts for issuing searches	• Issue all test searches using automated scripts • Identify any problems or issues • Revise and finalize automated scripts		Oct 1-15	Index Data Jason
Test programs for compiling results	• Review results of the searches and the reports generated • Identify revisions or enhancements needed • Finalize results reporting mechanisms		Oct 1-15	Index Data Jason
Documentation	• Develop summary documentation for validation and testing	Validation documentation	Oct 15-30	Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Testing Validation								

2.7 Analysis and Reporting

This work area addresses the development of standard documentation for reporting results of interoperability testing. Analysis procedures will be documented.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Analysis procedures	• Identify analysis procedures for results of testing • Test these procedures during 2.6 above • Finalize analysis procedures	Documentation of Analysis Procedures and Reports	Sept 15-Oct 1	Jason Penny
Report template	• Develop standard report template		Sept 15-Oct 1	Jason

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Analysis & Reporting								

2.8 Conduct Interoperability Testing with Libraries

This work area covers activities and tasks related to conducting interoperability testing with actual library online catalog and Z39.50 server implementations.

Activities	Tasks	Deliverables	Timeline	Responsibilities
Create testbed operation procedures	• Develop draft procedures document • Request review • Finalize procedures document	Testbed procedures document	Oct 1-15	Bill Jason
Radioactive MARC Record Availability	• Set up password protected web folder for records • Load records on site • Test access to records		Oct 1-15	Bill Jason
Limited interoperability testing	• Identify 3-5 libraries willing to go through testing procedure • Conduct Interop testing • Revise any procedures as necessary		Oct 15-Nov 30	Bill Jason Penny Index Data
Publicize testbed	• Prepare press release • Prepare call for participation • Disseminate information		Nov 15-Nov 30	Bill Jason Penny
Operation testbed	• Carry out interoperability testing with libraries		Nov 30-Feb 15	Bill Jason Penny Index Data

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
Interop Testing								

3. Project Timeline

The project currently has eight month duration. All work for the project should be completed by February 28, 2005. The letter approving the extension arrived in the second week of June. Substantial work on the project will begin on July 1, 2004 (allowing time for subcontracts to be issued and staffing to be confirmed). The following gives a high-level view of carrying out the project work.

Work Area	July04	Aug04	Sep04	Oct04	Nov04	Dec04	Jan05	Feb05
MARC DB								
MARC Records								
Test Searches								
Testing Scripts								
Indexing Policies								
Testing Validation								
Analysis & Reporting								
Interop Testing								

Attachment B: Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records

Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records

Introduction

In a first phase of the Z39.50 Interoperability Testbed, a large dataset of MARC records was used. In this work, we are exploring how a set of special, diagnostic MARC records can be developed and used to identify interoperability problems between a Z39.50 client and a Z39.50 server providing access to a database of bibliographic records supporting the search and retrieval functions of an online library catalog. We refer to these special, diagnostic records as radioactive MARC records. This document describes a framework for interoperability testing using these radioactive MARC records. It discusses the various components and identifies the tasks related to developing and implementing the components.

Search Requirements To Be Tested

The Z39.50 profiles for library application (Bath and the U.S. National) specify searches that Z39.50 clients and Z39.50 servers (and the underlying information retrieval system) should support if vendors' systems are conformant with the profiles. Each profile specifies levels of conformance, where each level requires the support of a set of searches. For this work, we will base our searches on the U.S. National Profile for Library Applications, approved by NISO as an American National Standard in 2003. Searches defined in conformance levels 0 and 1 will be tested. The following is a list of these searches. For details on the Z39.50 attribute combinations for each search, refer to the profile.

Level 1 Searches
Author Search – Keyword
Title Search – Keyword
Subject Search – Keyword
Any Search – Keyword

Level 2 Searches
Author Search – Keyword with Right Truncation
Author Search – Exact Match
Author Search – First Words in Field
Author Search – First Characters in Field
Title Search – Keyword with Right Truncation
Title Search – Exact Match
Title Search – First Words in Field
Title Search – First Characters in Field
Subject Search – Keyword with Right Truncation
Subject Search – Exact Match
Subject Search – First Words in Field
Subject Search – First Characters in Field
Any Search – Keyword with Right Truncation
Standard Identifier Search
ISBN Search
ISSN Search
Remote System Record Number Search
Date of Publication Search
Language Search
Format of Material Search – Keyword

These searches span a range of access points (e.g., author, title, subject, ISBN, ISSN, etc.), address precision oriented searches (e.g., exact match, first words in fields, first characters in fields) and recall

oriented searches (e.g., keyword, keyword with right truncation, phrase with right truncation), and searches that refine results (e.g., limiting by date, format of material, language).

Test Searches

The searching requirements will inform the development of test searches. One or more test searches may be necessary to adequately test the search types listed above. Each test search will define the type of search, the search term, and the expected results to be retrieved from the system being tested. Searches will be sent from the testbed client to ensure that appropriate Z39.50 attribute combinations are used to characterize the search term. The attribute combinations are defined in the U.S. National Profile.

The search term in the test searches will be aligned with specific data from specific fields/subfields in the radioactive MARC record. The retrieved records for each test search should be one or more of the radioactive MARC records, assuming the system being tested is in conformance with the U.S. National Profile.

The Radioactive MARC Records

One of the key research activities in this project is the development of a set of radioactive MARC records. This is particularly challenging because:

- MARC provides for rich content designation (e.g., various types of title information can be recorded in many different areas of the record)
- MARC records represent different formats of materials (e.g., books, computer files, etc.), and different formats may require the use of specific content designation not used in other formats

The number of required radioactive MARC records is not known at this time, but we anticipate creating at least one MARC record for each of the following formats of materials:

- Books, Pamphlets, and Printed Sheets
- Cartographic Materials
- Electronic Resources
- Continuing Resources
- Manuscripts (including manuscript collections)
- Music (Notated and manuscript music)
- Sound Recordings (musical and non-musical)
- Motion pictures and video-recordings (including digital and non-digital)
- Graphic materials (includes mixed materials, with or without archival control)
- Three Dimensional Artifacts and Realia

Multiple records for each format may be necessary to provide adequate coverage for the anticipated interoperability tests.

The MARC records will use the fields/subfields appropriate to each of these formats (since some MARC fields/subfields only apply to specific formats of materials). In each of the records, there will be patterns of alpha/numeric characters encoded in appropriate fields/subfields. The specifications for the character strings and how they will be generated to code into the records will be determined in discussions with project staff. An example of the concept follows:

The 245$a (title information) for two records might look like the following:

245 $a**Xysos sifys asdsai-aduus**
245 $a**Yszaqs sifys asdsai-aduus**

A title keyword search with the search term “sifys” should retrieve both of these records. A title exact match for the term “Xysos sifys asdsai-aduus” should only retrieve the first record.

Indexing Policies

During the first phase of the Z39.50 Interoperability Testbed Project, we developed indexing guidelines for several of the profile-defined searches (i.e., author, title, and subject keyword searches). When developing these guidelines, we identified approximately 500 subfields that could potentially store relevant data.

Since the testing planned for the current work encompasses searches beyond the keyword searches used in the first phase, it will be necessary to produce indexing guidelines for the additional searches. The indexing guidelines will be implemented in the Z-Interop Z39.50 server and underlying information retrieval system reference implementation. This is necessary to ensure baseline information about searching and retrieval behavior prior to conducting interoperability testing with other systems.

Testing Procedures

The intention of this work is to develop a set of automatic procedures to send the test searches and retrieve results. In the first phase of the project, sending the test searches was manually done by project staff. Automatic procedures can increase the efficiency of interoperability testing. These procedures should enable an increased number of test searches to be used. This will be important to assess the indexing policies in operation on the systems that will be tested. For example, we know that title information can be stored in different MARC subfields. The radioactive MARC records will have different patterns of characters in each of those subfields, and we can send individual searches with the specific string of characters that are stored in each of the subfields to see if the record is returned. If it is, we can conclude:

- The Z39.50 server is processing the search properly and sending it to the information retrieval system
- The information retrieval system does index the particular subfield.

The automated testing procedures will also assist in the management and analysis of results to reduce the manual effort that was required in the first phase of the testbed project. For example, the procedures should record any Z39.50 server diagnostic messages to understand if the server does not support particular types of searches; the procedures should record the search sent and the record retrieved; the procedures should enable compilation of results in an easy to use format for generating reports.

Availability of Radioactive MARC Records

For systems that choose to go through interoperability testing, the radioactive MARC records must be available for access and loading into the local system. The records will be stored on a password protected website from which system managers will be able to download the records. The records will be in conformant MARC 21 format, and system managers should be able to put the records into normal batch loading or other input queues.

Analysis and Reporting

Upon completion of interoperability testing, the results will be analyzed using manual and automatic methods. Wherever possible, automatic analysis should be used to streamline the process of analysis. Interoperability testing reports should include the following:

- Assessment of the level of Z39.50 interoperability
- Assessment of search functionality available

- Assessment of indexing policies.

Completed reports should be easy to understand by the managers of the system tested, and should enable them to take specific actions that might improve interoperability.

Tools and Resources Required

A number of tools will be required to set up and operate the testbed using the radioactive MARC records. The section describes anticipated tools and resources.

Database of MARC Content Designation

Because of the rich content designation available in MARC for bibliographic records, and because some of the content designation is intended for use when describing specific formats of materials, we will create a MARC Content Designation Database (MARC DB). This database will store information from the MARC 21 document and include the following:

- Field tags
- Indicator values, names, and definitions associated with field tags
- Subfield codes, names, and definitions associated with field tags
- Format applicable for each subfield code.

The MARC DB will enable queries to identify all appropriate content designation for the 10 formats of materials discussed above. In addition, the MARC DB will be used when generating reports of testing results to efficiently pull content designation names and definitions to make the reports easier to understand.
[Status: Under development by TxCDK]

MARC Record Creation Tool

To create the radioactive MARC records, we will use a basic MARC record creation tool or cataloging tool. The tool must allow the use of all available MARC content designation (either through a template or by editing). It would be best to use a tool that provides basic validation of MARC input. If possible, the tool should allow import from a CSV or other form of text file to get data into the record and then produce a conformant MARC 21 record.
[Status: To be identified and evaluated by TxCDK]

Radioactive MARC Records

A set of specialized, diagnostic MARC records will be created for use in interoperability testing.
[Status: Provided by TxCDK]

Z39.50 Client for Testing

A Z-Interop Z39.50 client will be used to send the test queries. The client must be conformant with Levels 0 and 1 of the U.S. National Profile.
[Status: Provided by Index Data?]

Z39.50 Server and Information Retrieval System Reference Implementation

To ensure the reliability of all aspects of the interoperability testing, we will use a reference implementation of a Z39.50 server and information retrieval system. Current plans are to use the existing Z-Interop reference implementation (the SIRSI Unicorn system). Current and new indexing guidelines will inform the indexing policies for the reference implementation. The radioactive MARC records will be loaded into the system and made available to searching. Searching and retrieval procedures will be tested and verified.

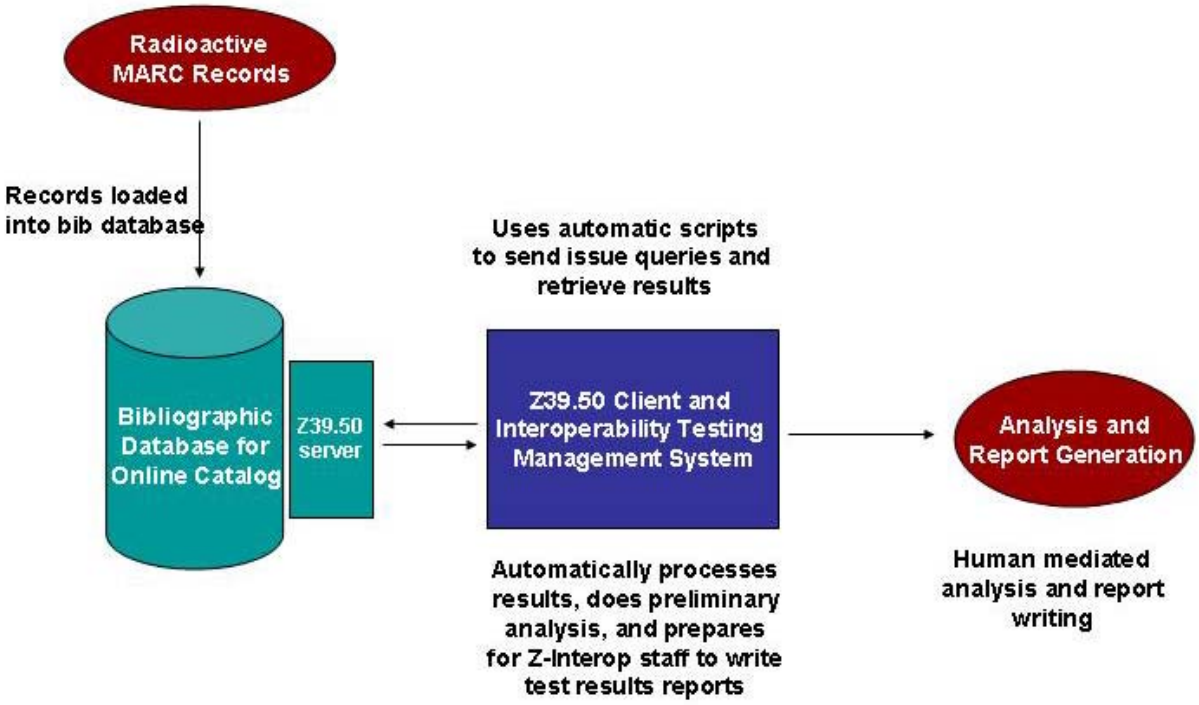
[Status: Probably needs some checking, but should be operational with a low level of effort]

Automatic Scripts for Testing and Reporting

Scripts will be used to send test searches from the Z-Interop client to systems.
[Status: Provided by Index Data?]

Visual Overview of Radioactive MARC Record Interoperability Testing

The basic interoperability testing framework can be graphically illustrated to highlight key components and processes involved as in the following.



Summary

This document is intended to initiate and guide the work for developing an approach for interoperability testing using radioactive MARC records. It has identified the components for interoperability testing and some of the tools and resources needed. Each of these components and processes will be described in more detail in separate documents. Additionally, a work plan document will identify the work areas, activities, and tasks to further guide the work and provide project management oversight for the work.