

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**

July 1 through December 31, 2004

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

January 1, 2005

Table of Contents

Introduction	2
Summary Accomplishments and Challenges	2
Project Personnel	3
Project Management	3
Project Website and Information Dissemination	3
Project Technology and Software	4
Discussion of Project Activities and Accomplishments	4
Summary and Next Steps	8

Introduction

This document serves as an interim report on the Z39.50 Interoperability Testbed Project, Phase 2 (Z-Interop2), covering the period of July 1, 2004 through December 31, 2004. 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 2004. This document highlights activities and accomplishments to communicate to IMLS progress on our project since the interim report on July 1, 2004.

Summary Accomplishments and Challenges

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

Accomplishments

- Designed and implemented the MARCdocs database
- Drafted several internal documents including:
 - *Z-Interop2 Project Search & Record Data Requirements for Z39.50 Interoperability Testing Using Radioactive MARC Records*
 - *Specifying Tokens for Radioactive MARC Records to Support Interoperability Testing of Selected Searches*
 - *Radioactive MARC Records Specifications*
 - *Z-Interop 2 Project Test Searches Using Radioactive MARC Records*
- Created first set of radioactive MARC records according to draft specifications
- Successfully conducted proof of concept interoperability testing using radioactive MARC records
- Developed and tested software (Index Data) to automatically send test searches, gather results, and produce reports.

Future Challenges

- Refine radioactive MARC record creation according to revised specifications and requirements
- Identify potential libraries to participate in interoperability testing using the radioactive MARC records and project testing software
- Determine a reasonable set of radioactive MARC records that will be sufficiently robust to support meaningful interoperability testing.

The activities accomplished to date have provided a proof of concept for this approach to interoperability testing. However, discussions on the specifications for the radioactive MARC records have taken more time than originally envisioned. This combined with the approval from IMLS for this work coming several months later than had been originally proposed in the request for extension submitted in February 2004 makes it difficult to successfully complete all work for the project by the end of February 2005. The PI, in a

separate communication to IMLS, will request a no-cost extension of 6 months for this project.

Project Personnel

The project team consists of:

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

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

- Penelope Benardino, a recent graduate from the School of Library and Information Sciences who worked on the Z-Interop Project in 2003. She completed her contract in November 2004.
- Index Data, a software firm specializing in Z39.50 that was listed in the request for extension to provide technical support for the project. Index Data is continuing its work through the end of the project.

Each of the project team and contractors has been assigned activities and tasks related to project goals and objectives.

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 produced several internal draft documents related to work on the project. These documents are discussed below in more detail. These documents reflect the PI's responsibility for the design and conceptualization of the project. The contractors have played a valuable role in reviewing these documents and providing suggestions for improvement.

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 has yet to be modified to reflect the work of Phase 2 of the Z-Interop Project, but this is listed as a priority task for the month of January 2005. The PI plans to submit announcements of the Phase 2 work to alert the community of this work and to solicit a limited number of libraries to participate in interoperability testing during the project.

Project Technology and Software

In addition to desktop PCs used by the project team (at UNT), the project is also using one Linux server that hosts the MARCdocs database, and the Z-Interop testbed server that was implemented as part of the project's Phase 1. No IMLS funds were used for any technology or software. The Z-Interop2 Project is leveraging many of the hardware and software resources acquired during the original Z-Interop Project.

Discussion of Project Activities and Accomplishments

There were four primary areas of work in which the project team and contractors engaged during this reporting period:

- Creation of a database of MARC documentation
- Identification of searching and data requirements for interoperability testing
- Specification for and creation of the radioactive MARC records
- Development and testing of software to support interoperability testing.

Each of these is discussed below.

Creation of a Database of MARC Documentation

One of the challenges of creating a sustainable interoperability testing environment is to identify potential components of a testbed that can support the automation of activities and procedures. One assumption for the project was that automatic testing software, developed by Index Data, could formulate appropriate test searches, issue those to specific search targets, gather results, and produce reports. We identified early in the project that a database would be a critical component for the testbed.

Based on the requirements for the test searches, we determined that a database of MARC documentation would serve multiple purposes. This would allow the storage of information about all content designation available in the MARC 21 Format for Bibliographic Data, and such information could assist in the creation of the radioactive MARC records. In addition, a flexible and extensible relational database would allow the storage of information about profile-defined searches necessary to the automatic testing software.

The result of this work is the MARCdocs database available in a public version at: <http://meta.lis.unt.edu/MARCdocs2/>. The working version for the Z-Interop2 Project is not publicly accessible. Surprisingly, there is not a publicly available database of MARC documentation, and we have announced this first release of the MARCdocs database to the broader MARC community (and clearly crediting support from IMLS for its creation). Further, the PI received a new IMLS National Leadership Grant to continue his investigation of MARC content designation utilization began as part of the original Z-Interop Project. This MARCdocs database will be a critical tool for that project, and again, leveraging IMLS support for ongoing and new research.

The MARCdocs database has gone through several iterations, and continues to evolve as we understand how this tool can be used to support the Z-Interop2 testbed approach. Complete documentation for the MARCdocs database and its applications will be finalized and mounted on the project website for public consumption in the first quarter 2005.

Identification of Searching and Data Requirements for Interoperability Testing

During June 2004 as Z-Interop2 work was initiated, the PI developed *Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records*, a conceptual framework that identified the primary components of the radioactive MARC records approach for interoperability testing. Two components identified in the framework document were further elaborated in an internal project document, *Z-Interop2 Project Search & Record Data Requirements for Z39.50 Interoperability Testing Using Radioactive MARC Records*. This document analyzed the data requirements for the radioactive MARC records in the context of the questions the testing attempts to address.

After some deliberation, we identified four different levels, each with a different focus, that questions about interoperability could be asked and answered. While there may be other levels, these provide the point of departure for the Z-Interop2 work:

- **Profile conformance level:** This level addresses the interoperability between the Z-client and Z-server. Assessing this level of interoperability relies on the use of Z39.50 profiles that identify Z39.50 specifications for search and retrieval. Questions that can be addressed at this level include:
 - Does the Z-server process each query successfully?
 - If the Z-server cannot process the query as sent, does it send the appropriate diagnostic message?
- **Information retrieval system level:** This level addresses the capability of the information retrieval (IR) system underlying the online catalog application. Questions that can be addressed at this level include:
 - What search functionality does the information retrieval system have?
- **Bibliographic record level:** This level is also an IR system oriented level, but its focus is on how the IR system indexes fields in the bibliographic record to provide access points or searchable components of the record. Questions address by this level include:
 - Does the information retrieval system index the appropriate fields in the records for specific access points?
 - Do the system's indexing policies support searches for the searches defined in the Z39.50 profile?
- **Data content level:** This level takes us into more difficult arenas, but arenas that may affect interoperability. Here we are addressing how the IR system processes the data content of the records, such as questions related to normalization of the data, dealing with hyphenated works, and special characters and diacritics. Although we identified this as a level at which questions could be

asked and answered, the resources available in this project do not allow the exploration of this level.

This document provided the first concrete set of data requirements for the radioactive MARC records. The document also identified a set of profile-defined searches that would be tested in the Z-Interop2 testbed. It also led to additional discussions among the project team and contractors on the details of the radioactive MARC records. The results of which are discussed next.

Specification for and Creation of the Radioactive MARC Records

The specially constructed MARC records for this approach to interoperability testing are the foundation, and the design of these records was a key intellectual challenge for the team. In a set of internal draft documents and many email exchanges, we worked through ideas for what data values would be contained in the radioactive MARC records that would support interoperability testing. The result was the draft document, *Radioactive MARC Records Specifications*. The following gives some information about the specifications.

The fundamental data unit in the radioactive MARC records is a token. A token is a string of characters that has a specific structure and semantics that will serve as “words” or other data values in specific fields/subfields. A field/subfield may have a sequence of tokens. The specially designed tokens will populate each indexable field. The following structure results in content-rich token elements:

- A single alpha character for left-hand padding.
 - Value = r
- A single alpha character to indicate the format of the material being described or type of record
 - Value = Selected values as defined in MARC Leader/06 – Type of Record or the Leader/07 – Bibliographic Level
- Three numbers indicating the Field Tag
 - Value = Defined in MARC 21 specifications
- A single integer to indicate number of occurrence the Field Tag
 - Value = Sequential number starting with 1
- A single alpha character to indicate the Subfield Code
 - Value = Defined in MARC 21 specifications
- A single integer indicating the offset within subfield
 - Value = Use the following scheme: 1=first token in subfield, 2=second token in subfield; 3= third token in subfield, etc.
- A single alpha character for right-hand padding
 - Value = r

And example token element to show this structure is: **ra2451a1r**. We can parse it as:

- r - Left-hand padding

- a - Type of record -- this is a books type record
- 245 - Field code
- 1 – First occurrence of field in record
- a - Subfield code
- 1 - Offset within subfield -- 1=first token in subfield
- r - Right-hand padding

In addition to specifying the structure of the tokens, other information needs to be encoded in the radioactive MARC record, including information to uniquely identify the record, the version of the record, and other details about the source and purpose of the record.

The following is an example of a human-readable radioactive MARC record built according to the specifications

001	UNTRadMARC001
040	\$a ZinteropUNT
100	\$a rm1001a1r, rm1001a2r, \$d rm1001d1r.
245	\$a rm2451a1r rm2451a2r rm2451a3r : \$b rm2451b1r rm2451b2r rm2451b3r / \$c rm2451c1r rm2451c2r rm2451c3r.
440	\$a rm4401a1r rm4401a2r rm4401a3r
490	\$a rm4901a1r rm4901a2r rm4901a3r
583	\$a RadMARC \$b www.unt.edu/zinterop/001 \$d 1 \$e ATS \$i 1 \$k JungWon Yoon \$x This is a specially created test record for the Z-Interop2 Project under the direction of the Texas Center for Digital Knowledge at the University of North Texas. Contact Dr. William E. Moen via email for information about this project at <wemoen@unt.edu>. Funding for this project is provided by the U.S. Federal Institute of Museum and Library Services. This particular record supports testing related to a Books, Pamphlets, and Printed Sheets type of record that was created to support test searches for author, title, subject, and any profile-defined searches, where the threshold of occurrence of the indexable content designation being populated in the record is 19, for the 19 most commonly occurring indexable author, title, and subject fields. This is the first version of this record.
600	\$a rm6001a1r rm6001a2r, \$d rm6001a1r.
650	\$a rm6501a1r rm6501a2r rm6501a3r \$x rm6501x1r \$v rm6501v1r \$z rm6501z1r.
651	\$a rm6511a1r rm6511a2r \$x rm6511x1r.
653	\$a rm6531a1r rm6531a2r rm6531a3r
700	\$a rm7001a1r rm7001a2r, \$d rm7001d1r.
710	\$a rm7101a1r rm7101a2r.

A set of 15 radioactive MARC records was created based on an earlier version of the specifications document to allow early testing within the Z-Interop testbed reference implementation Z39.50 server and online catalog database. This enabled the project team to prove out the concepts and specifications developed to date in the project. Through this validation of procedures and specifications we confirmed our approach at both the conceptual and technical levels.

Development and Testing of Software to Support Interoperability Testing

The final area of work was done by Index Data, the contractor on the project. Index Data is responsible for developing the automatic testing software and scripts to enable the following:

- Identify a specific profile-defined test search to formulate
- Formulate the test Z39.50 query to conform to the profile-defined search specification
- Determine the appropriate search term in the form of one or more tokens for the test search
- Send the test search to a search target
- Retrieve results of the search
- Analyze the results of the search
- Produce a report with the results of a set of test searches

To date, Index Data has developed the software and tested it successfully against the Z-Interop testbed reference implementation Z39.50 server and online catalog database. The reference implementation online catalog database contained the preliminary set of radioactive MARC records.

Summary and Next Steps

We have made great progress on this project, and most importantly, we have preliminary validation for the conceptual and technical approach to interoperability testing using radioactive MARC records, the MARCdocs database, and the automatic testing software produced by Index Data.

With this foundation, our next steps are the following:

- Create appropriate sets of radioactive MARC records for use in interoperability testing
- Validate the radioactive MARC records for correctness
- Validate the Z-Interop reference implementation containing the radioactive MARC records for proper behavior
- Retest the automatic testing software against the Z-Interop reference implementation to validate all scripts and reports
- Identify 3 to 5 libraries that are willing to load the radioactive MARC records and have interoperability testing conducted on their implementations
- Complete all documentation for the components of the testbed
- Update the Z-Interop Project website to make publicize the Z-Interop2 Project
- Write one or more articles for publication about this project and its approach to interoperability testing.

Creating the sets of radioactive MARC records and soliciting participation of 3-5 libraries will likely be the most time consuming. To ensure sufficient time is available for successful completion of this project, we will request a no-cost extension from IMLS to continue this work for another six months.