PREPRINT COPY – DO NOT COPY OR CITE

Accepted for Presentation at the 2005 ASIST Annual Meeting

An Extensible Approach to Interoperability Testing: The Use of Special Diagnostic Records in the Context of Z39.50 and Online Library Catalogs

William E. Moen, Ph.D.

Texas Center for Digital Knowledge, University of North Texas
<wemoen@unt.edu>

Sebastian Hammer & Mike Taylor

Index Data, Copenhagen, Denmark

Jason Thomale & JungWon Yoon

Texas Center for Digital Knowledge, University of North Texas

Assessing interoperability in the networked information services and applications environment presents difficult challenges due in part to the multi-level and multi-faceted aspects of interoperability. Recent research to establish an interoperability testbed in the context of Z39.50 protocol clients and servers and online catalog applications identified threats to interoperability and defined a question space for interoperability testing. This paper reports on follow-up research to develop an alternative approach for interoperability testing in the context of networked information retrieval that uses specially designed diagnostic records. These records, referred to as radioactive records, enable interoperability assessment at the protocol and semantic levels. This approach appears to offer an extensible method for interoperability testing for other metadata and protocol application environments. The resulting interoperability testbed incorporates additional components to exploit automatic processes for interoperability testing and assessment, thus improving the efficiency of interoperability testing.

Introduction

Pursuit of interoperability in the networked information environment has been compared to the pursuit for the Holy Grail (Tennant, 1998). We believe it exists, and we believe we can find it (or achieve it). Testing for interoperability in basic networked services and applications such as information retrieval have often resembled the Keystone Kops in the simplicity of some approaches, or Rube Goldberg machines for conformance testing – far from the sublime pursuit of the Grail. Yet the challenges in achieving useful levels of interoperability are problematic in part because of the multi-faceted nature and types of interoperability (Miller, 2000).

One networked information service area that has provided an opportunity to explore the multi-faceted nature of interoperability is the use of the Z39.50 information retrieval protocol to conduct information retrieval tasks on a variety of online databases, including bibliographic databases associated with online catalog applications. We have seen how optimal interoperability must occur not only at the syntactic or functional level provided by the protocol but also at the semantic level. This latter level addresses ability of two systems to present and process user information tasks in a way that meanings of those tasks are retained. Reliability, trustworthiness, and usability of networked resources and services are founded on assumptions about the levels of interoperability occurring when two or more systems interact in service to applications and users.

This paper describes an on-going research project to explore issues related to interoperability in the context of metasearch applications across multiple online library catalogs or bibliographic databases. The immediate goal of this research is to improve interoperability when using the Z39.50 information retrieval protocol. The paper presents a new approach to interoperability testing through the use of specially-designed Machine-Readable Cataloging (MARC) records, which we call radioactive MARC (RadMARC) records.

Background

The literature on the topic of interoperability is both broad and deep, and continues to expand. From brief overviews describing interoperability (see Miller, 2000) to more technical treatments (see Lynch & Garcia-Molina, 1995) to the implications of interoperability on policy (see Moen, 2001a), the literature treats interoperability from multiple perspectives.

The U.S. Federal Institute of Museum and Library Services (IMLS) awarded a National Leadership Grant to the Texas Center for Digital Knowledge in 2000 for a research project to explore the issues of interoperability among online library catalogs and their bibliographic databases accessible via the Z39.50 protocol. The overall goal of the Z39.50 Interoperability Testbed (Z-Interop) Project was to improve Z39.50 semantic interoperability among libraries for information access and resource sharing. Information about this phase of the research, including the full proposal and various reports, is available at: <http://www.unt.edu/zinterop>.

Several key components of the testbed included:

- **Test dataset:** Approximately 400,000 MARC 21 bibliographic records from OCLC's WorldCat database.
- **Reference implementations:** Reference implementations of a Z39.50 server and an information retrieval system (in the form of an online catalog) using an integrated library system from Sirsi Corporation. Reference implementation for a Z39.50 client using the Bookwhere Z39.50 client. The reference implementations were under the control of the Z-Interop project staff to configure according to published specifications in the form of Z39.50 profiles.
- **Test searches and benchmarks:** A set of test searches as defined in the Z39.50 profiles and benchmark results for each test search established by using the reference implementations.

Project staff assumed that the target audience for interoperability testing would be the vendors of Z39.50 client and server products, and individual libraries that wanted to check their systems interoperability with the reference implementations. Although the project was successful in recruiting participation from vendors of products, we discovered that individual libraries did not have the capability to load the 400K test dataset into their systems. Hardly any libraries have a test environment for their implementations, and this was a basic limitation of the Z-Interop Project's testbed approach. However, vendors of both Z39.50 servers and clients went through interoperability testing. The testbed proved fruitful in better understanding several factors that affect interoperability, and also demonstrated that with some attention by the vendors, their products could be configured to achieve 100% interoperability using the testing procedures provided by the testbed.

However, the true arena for interoperability testing is not just the vendors' products but their actual instantiation in a particular implementation, namely the implementation of the product as a production-level application in a library.

An Alternative Approach for Interoperability Testing

The limitations of the testbed approach described above motivated an investigation for an alternative method for interoperability testing for Z39.50 servers and library bibliographic databases. IMLS provided additional funding to continue the Z-Interop Testbed to explore this alternative approach.

The alternative method uses a small set of very special MARC records (we refer to these as "radioactive MARC records," explained below) that can serve as diagnostic mechanisms for assessing system functionality, performance, and interoperability. This alternative approach has potential for providing interoperability testing services to individual libraries. In addition, this approach may be adaptable to other protocol and metadata contexts beyond Z39.50 and MARC.

The metaphor of a radioactive MARC record is based on current medical diagnostic techniques for people. When a person has a particular medical condition, there may be two approaches for diagnosis. One could be considered invasive, where the person would undergo a surgical technique for physically examination of the problematic area or anomaly. The other approach could be considered less invasive, where the patient is injected with a dye, possibly radioactive, and once it has spread throughout the body, scanning techniques allow a medical professional to identify structural or mechanical problems or anomalies.

A "radioactive" MARC record approach for interoperability testing is less "invasive" for an individual library. It does not require loading a large test dataset such as used in the first Z-Interop testbed. Nor does it require a separate testing environment on the local implementation. Instead, the library loads these special MARC records into its production online catalog system, and the Z-Interop staff conducts a series of tests to assess system functionality, performance, and interoperability. The radioactive MARC records are legitimate instances of MARC records that a library system can import and process, and then remove when the testing is completed. These records, however, have very special characteristics.

The Threats to Interoperability

In the first phase of the Z-Interop testbed, we anticipated several levels at which interoperability needs to occur, and we identified some of the threats to such interoperability. The research in that project confirmed the reality of these threats.

In a broader context, Moen (2001b) identified a number of diverse factors that can affect interoperability in networked information retrieval applications:

- Multiple and disparate operating and Information retrieval systems
- Multiple protocols
- Multiple metadata schemes
- Multiple data formats
- Multiple languages and character sets
- Multiple vocabularies, ontologies, and disciplines.

In the context of the Z-Interop Project, we identified key factors threatening interoperability:

- Differences in implementation of the standard
- Differences in local information retrieval systems

In the latter case, this includes search functionality available in the system, indexing policies affecting the access points in the database, word extraction and processing choices, and character set and character encoding and normalization. As a way to indicate the scope of our investigations, Moen (2001a) identified the levels of interoperability of concern in the Z39.50 context:

- Low-level protocol (syntactic): Do Z-client and Z-servers interchange protocol messages according to the standard?
- High-level protocol (functional): Do Z-client and Z-servers support appropriate Z39.50 information retrieval services for user tasks?
- Semantic level: Can Z-clients and Z-servers and local information retrieval systems preserve and act on meaning of information retrieval tasks?
- User Task level: Do systems support information retrieval tasks of one or more user groups?

Within the context of our investigations and the maturity of Z39.50, the syntactic level is of little concern. The development of the Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery (2004), and the U.S. National Z39.50 Profile for Library Applications (National Information Standards Organization, 2003) addressed many issues related to the functional level. The Z-Interop testbed was successful in part because the profiles defined expected Z-client and Z-server behaviors and interactions. The biggest challenge to reliable interoperability appears at the semantic level. Semantic interoperability here is not addressing the concerns of two words meaning the same thing or other problems related to linguistics and meaning. Instead, semantic interoperability concerns the ability of two systems to present and process user information tasks in a way that meanings of those tasks are retained. For example, if a user does a title search for information resources, the search is actually executed on a search target against words from titles in the record. A common sense idea, yet often search targets do not process searches as the user intended (e.g., processing an exact match search for a title as a set of keywords combined using Boolean operators and matching the words not only in title access points but in other access points as well).

The Question Space for Interoperability Testing

In our alternative approach for interoperability testing we identified a set of questions that could be asked to address the different levels of interoperability. These questions pointed to appropriate test searches and the data needed in the records. The following summarizes the question space for our interoperability testing:

- **Profile conformance 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 (IR) system level:** Addresses the capability of the IR system underlying the online catalog application. Questions that can be addressed at this level include:
    - Does the IR system have the requisite search functionality to support the searches defined in the Z39.50 profiles?
- **Metadata record level:** Also an IR system focus, but concerned with how the IR system indexes fields in the metadata 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:** Addresses how the IR system processes the data content of the records, such as processes related to normalization of the data, dealing with hyphenated works, and special characters and diacritics.

The question space is also informed by two Z39.50 profiles:

- ANSO/NISO Z39.89, The U.S. National Z39.50 Profile for Library Applications (National Information Standards Organization, 2003) <http://www.niso.org/standards/resources/Z39_89final.pdf>
- Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery, Release 2.0 (The Bath Group, 2004)

These specifications provide well-defined searches and expected client and server behaviors at several conformance levels. For initial interoperability testing, we used the profile-defined searches listed Table 1. These searches also pointed to the data necessary in the RadMARC records to support the testing using these searches.

Table 1. Z-Interop Testbed Search Types Approximately Here

Components of the Radioactive MARC Record Interoperability Testbed

In addition to the reference implementations used in the original Z-Interop testbed, the alternative testing approach introduces three new components:

- The specially designed MARC records
- A set of test searches and automatic testing script that issues searches, retrieves records, and develops reports on the search and retrieval results
- A database of MARC documentation that enables the automatic identification of types of searches to issue.

The basic interoperability testing framework is illustrated in Figure 1 to highlight key components and processes.

Figure 1. Interoperability Testing Framework Approximately Here

Radioactive MARC Records

As noted before, two Z390.50 profiles provide the specifications that are the basis for the test searches. Searches defined at Level 0 and Level 1 require appropriate RadMARC records for the test searches. In addition, different types of records (as indicated in the MARC Leader/06) are needed since systems may index MARC fields/subfields differently depending on the type of record. Finally, the RadMARC records need to be clearly identified as to the type of searching they are intended to assess, and therefore some version information about each record is included in the record.

MARC records can describe different types and formats of bibliographic materials. MARC Leader/06 indicates the type of record (i.e., the information object type being described by the MARC record). The *Anglo-American Cataloguing Rules* (AACR), the standard for descriptive cataloging, specifies rules for describing different types of materials. MARC Leader/07 indicates the bibliographic level of the record. Table 2 summarizes the values of the MARC Leader/06, Leader/07, and the AACR categories of materials to show the complexity of coding and labeling for what the MARC records describe.

The coding of Type of Record in MARC Leader/06 is not aligned directly with the 10 format types of information objects as addressed by AACR (in the third column of the table above). In some cases, two code values in the Leader/06 are addressed by the same AACR format of material. Additionally, the

Leader/06 doesn't indicate if the material is a serial publication (or Continuing Resource). For this, the Leader/07 – Bibliographic Level indicates serial with a value of s. To summarize, the types of records or materials for which RadMARC records were created address those in the table.

Table 2. Types of Materials Described by MARC Bibliographic Records

Approximately Here

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. The fundamental data unit in the RadMARC 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 populate selected field/subfields in the RadMARC records. Several sets of RadMARC records are used in interoperability testing. The sets are distinguished by the amount of content designation populated in the records (see discussion below). All selected content designation use the special tokens. The following is the structure of content-rich tokens being used in the RadMARC records:

- A single alpha character for left-hand padding.
  - o Value = r
- A single alpha character to indicate the format of the material being described or type of record
  - o Value = Selected values as defined in MARC Leader/06 – Type of Record or the Leader/07 – Bibliographic Level
- Three numbers indicating the Field Tag
  - o Value = Defined in MARC 21 specifications
- A single integer to indicate number of occurrence the Field Tag
  - o Value = Sequential number starting with 1
- A single alpha character to indicate the Subfield Code
  - o Value = Defined in MARC 21 specifications
- A single integer indicating the offset within subfield
  - o 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
  - o Value = r

An example token that shows this structure is **ra2451a1r**, which can be parsed 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, where 1 = first token in subfield
- r - Right-hand padding

In addition to the field- and subfield-specific tokens, each RadMARC record contains additional 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 RadMARC record in human-readable form built according to the specifications.

Figure 2. Sample RadMARC Record

Approximately Here

Automated Testing Scripts and Processes

Once a server's database has been injected with one or more radioactive records, a client can test that server's indexing and searching functionality by issuing searches that expect to return specific records. For example, a server that contains a record with a particular token in its 245$a should yield the record when queried with a search for that token against a title index, and using the appropriate Z39.50 query to express the query (e.g., as defined by Bath profile "title keyword" query). Conversely, so long as the same token does not appear elsewhere in the record, a search for that token in a subject index should *not* find the record.

Test searches such as these may be sent by any conforming Z39.50 client, but it is more efficient to automate testing using ready-rolled scripts. We took a two-level approach to building such scripts: at the low level, we created a domain-specific "little language" specialized for such scripts; and at the higher level, we created an initial set of scripts in that language, both as a useful partial test-suite for servers claiming Z39.50 profile conformance, and as proof of concept of the language/script division.

Although initial designs for the scripting language consisted of only a few domain-specific primitives, it quickly became apparent that scripts may in general need to make use of logical and looping constructs, and perhaps variable assignments and procedure definition/invocation, such as are provided by mainstream programming languages. Accordingly, we decided that the most efficient approach would be to build our language on top of a well-supported, expressive, existing language. Practical considerations indicated that Perl was the most appropriate choice, although Python would have been an attractive alternative were it better appreciated and more widely adopted in the Z39.50 community.

Our strategy, then, was to extend Perl with a “RadioMARC” module to allow Z39.50 searching of servers known to contain copies of specific records, and to emit reports dependent on whether or not the expected records are present in the result set. Perl's own language constructs are used in more complex test scripts to determine at run-time which tests to attempt, depending on the results of earlier tests.

A typical simple script follows:

```
use Net::Z3950::RadioMARC;
set host => 'z3950.loc.gov', port =>
'7090', db => 'voyager';
set delay => 3;
add "filename.marc";
test '@attr 1=4 01245a01', { ok =>
'245$a is searchable as 1=4',
notfound => 'This server is broken'
};
```

This illustrates the three important domain-specific operations, **set**, **add** and **test**. Once the RadioMARC module has been introduced (the “use” statement on the first line), these may be freely used:

- **set** merely sets the value of named parameters – in this case, the connection details for the server to be tested, and the number of seconds to delay between searches in order to avoid overloading the server.
- **add** registers a set of MARC records, added from the named file, which are believed to exist in the server being tested.
- **test** does the real work. First, it creates the Z39.50 connection if no connection has already been forged. Then it performs the search specified as its first argument. This argument expresses the query in the widely used Prefix Query Format (PQF), as described in the *YAZ User's Guide and Reference* (Hammer, et al., 2004), The same query is used on the client side to select which of the previously **add**ed records is the target for the query, and the result set returned by the server is inspected for that record's presence. A message is emitted depending on whether or the record is found, or whether the search failed completely – for example, because the server does not support the specified access-point.

As part of the deliverables from this research project, the RadioMARC Perl module will be released for public use.

MARC Documentation Database

One of the challenges of creating a sustainable interoperability testing environment is to identify potential components of a testbed than can support the automation of activities and procedures. The previous section discussed one aspect for automatic testing software that formulates appropriate test searches, issues those to specific search targets, gathers results, and produces reports. To support those functions we developed a database of MARC documentation that would serve multiple purposes

The database stores information about all content designation available in the MARC 21 Format for Bibliographic Data specifications. In addition, the flexible and extensible structure of the underlying relational database allows the storage of information about profile-defined searches necessary to the automatic testing software. Further, we examined how the database could assist in the creation of the radioactive MARC records.

MARCdocs: The MARC 21 Documentation Database, is a pilot effort aimed at structuring the textual documentation from the MARC 21 Format for Bibliographic Data into a relational database. Using a database approach for authoritative MARC documentation provides new opportunities for various applications. This database application uses open source software tools including Linux, MySQL, and PHP. A public version of the application is available at: <http://meta.lis.unt.edu/MARCdocs2/>. A working version of MARCdocs for our current research contains additional project-specific information and is not publicly accessible. Figure 3 is a screen shot showing example MARC field information in the database. Having this documentation, along with other project-specific data included in structured format in the database provides opportunities for automating many aspects of the testbed.

Figure 3. MARCdocs Database Interface Approximately Here

The RadMARC Content

In the discussion above about the RadMARC records, we indicated that the need to know what content designation in a MARC record to populate with tokens to support the interoperability testing of profile-defined searches. This is tied closely with a specific level of the question space for this interoperability testbed, namely:

- **Metadata record level:** This level focuses on how the information retrieval system indexes fields in the metadata 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?

As part of the original Z-Interop testbed, we identified more than 500 MARC fields/subfields in a MARC record that could be indexed to support author, title, and subject searching. The complete list of this content designation is contained in *Indexing Guidelines to Support Z39.50 Profile Searches* (Moen, 2002). In another analysis for that project, we analyzed occurrences of MARC content designation in the Z-Interop test dataset of more than 400,000 MARC records from OCLC's WorldCat database (Moen and Benardino, 2003). We also examined the occurrence of content designation that could be indexed to support author, title, and subject searches, and discovered that 19 of the more than 500 subfields that could be indexed accounted for 80% of all occurrences. Table 3 shows the top 5 of these 19 subfields.

Table 3. Top 5 MARC Indexable Subfields Approximately Here

The data resulting from that analysis were added to the MARCdocs database, which enables us query the database and identify content designation that could be indexed to support the Z39.50 profiles Level 0 and Level 1 searches. We use those frequency counts to select sets of fields/subfields to populate various sets of RadMARC records. The MARCdocs database, then, is used in the creation of the RadMARC records by holding information that designates a specific content designation as a candidate for indexing for particular searches and the frequency count of its occurrence based on the earlier analysis.

RadMARC Record Sets

We have identified several possible sets of RadMARC records to create and have completed the creation of two of these. The first set uses the 19 most commonly occurring indexable fields for author, title, and subject-related data discussed previously. The RadMARC record in Figure 2 shows how these content designation structures have been populated.

The second set of RadMARC records uses all author, title, and subject content designation that occurred 1,000 or more times from the earlier analysis. However, we can extend this to include all possible content designation as listed in the Z-Interop *Indexing Guidelines* document.

Two other sources of information are informing the creation of a third set of RadMARC records:

- The Network Development and MARC Standards Office (n.d.) recommendations for national level records
- The Program for Cooperative Cataloging (2003) core record standards.

For example, the recommendations for national level records identify "mandatory" and "mandatory if applicable" content designation. A comparison of those recommended content designation structures with the Z-Interop indexing guidelines indicate 131 fields/subfields that are author, title, and subject related. We can create RadMARC records using these 131 fields/subfields.

We think that the RadMARC approach can be used to develop any set of RadMARC records as well as custom built diagnostic records libraries can use to interrogate their systems' behavior. We are currently in discussions with a number of libraries that want to diagnose the indexing policies actually in effect on their systems to verify vendor configurations. We can create individual RadMARC record that are intended to exercise specific indexing policies.

The Extensibility of the Radioactive Record Approach

To date (Spring 2005), we have proved out the concept and the technologies involved with the RadMARC approach to interoperability testing using Z39.50 clients and servers and online catalog applications. We have identified additional sets of RadMARC records that can be created to analyze more deeply information retrieval system search functionality and indexing policies. The radioactive record approach, however, has the potential to be used to diagnose other system behaviors, for use in other metadata environments, and for use in other protocol environments. We discuss three potential uses below.

In Section 5 we described the question space for the interoperability testbed. The data content level is not being addressed in the current testbed because of resource constraints. However, it will be possible to create RadMARC records where the tokens include special characters and diacritics, which can then show how a local IR system normalizes or otherwise processes the data that may affect search results.

Another opportunity is in the Open Archives Initiative (OAI) metadata harvesting environment. One of the issues that OAI data providers face is mapping a rich native metadata scheme to simple Dublin Core metadata elements. In the case where library catalogs are being harvested, it would be interesting to see the effects of such local mapping decisions when the source records are specially designed RadMARC records.

Still another opportunity is to explore the creation of RadDC records (radioactive Dublin Core) or radioactive records using other metadata schemes to assist in diagnosing and making visible system behaviors in those application areas.

Finally, new protocols such as Search and Retrieve for the Web (SRW/SRU) could benefit from the current work by providing RadMARC records (or records in other metadata schemes) for information retrieval systems accessible by SRW. For example, a SRW search of a database of bibliographic records (possibly in MARC format) with a request to return results as DC records would allow diagnosis of various system behaviors (search access, search functionality, mapping from native database scheme to DC).

Summary and Conclusion

The current research project is establishing an innovative conceptual and technical foundation for interoperability testing. The scope of this research focuses currently on Z39.50 and online catalogs. The project has provided the opportunity to conduct proof-of-concept for a radioactive record approach for diagnosing interoperability factors in an identified question space. We see this approach as extensible in terms of the current focus (i.e., being able to create different sets of RadMARC records to diagnose general or specific interoperability issues) and to other application environments, metadata schemes, and protocols.

ACKNOWLEDGMENTS

The research for this project has been supported by a National Leadership Grant from the U.S. Federal Institute of Museum and Library Services.

REFERENCES

The Bath Group. 2004. The Bath Profile: An International Z39.50 Specification for Library Applications and Resource Discovery, Release 2.0. Available URL: <http://www.collectionscanada.ca/bath/tp-bath2-e.htm>

Hammer S. et al. 2004. YAZ User's Guide and Reference. Available URL:

<http://www.indexdata.dk/yaz/doc/>

Lynch, C. & Garcia-Molina, H. 1995. Interoperability, scaling, and the digital libraries research agenda: A report on the May 18-19, 1995 IITA Digital Libraries Workshop. Available URL: <http://www-diglib.stanford.edu/diglib/pub/reports/iita-dlw/>

Miller, P. 2000. “Interoperability: What it is and why should I want it.” Ariadne, 24. Available URL: <http://www.ariadne.ac.uk/issue24/interoperability/intro.html>

Moen, W.E. and Benardino, P. 2003. “Assessing metadata utilization: An analysis of MARC content designation use.” In 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice – Metadata Research and Application. Seattle, WA, September 28-October 2, 2003. Seattle: Information School of the University of Washington, 2003.

Moen, W.E. 2002. Indexing Guidelines to Support Z39.50 Profile Searches. Texas Center for Digital Knowledge, University of North Texas. Available URL: <http://www.unt.edu/zinterop/Documents/IndexingGuidelines1Feb2002.pdf>

Moen, W.E. 2001a. “Assessing interoperability in the networked environment: Standards, evaluation, and testbeds in the context of Z39.50.” In Evaluating Networked Information Services: Techniques, Policy, and Issues (pp.85-109), edited by Charles R. McClure and John Carlo Bertot, Publisher: ASIS thru Information Today, Inc. 2001.

Moen, W.E. 2001b. "Mapping the Interoperability Landscape for Networked Information Retrieval." In Proceedings of First ACM/IEEE-CS Joint Conference on Digital Libraries, Roanoke, VA, June 24-28, 2001 (pp. 50-52). New York: The Association for Computing Machinery.

National Information Standards Organization. 2003. ANSI/NISO Z39.89 - 2003 The U.S. National Z39.50 Profile for Library Applications. Bethesda, MD: National Information Standards Organization. Available URL: <<http://www.niso.org/standards/resources/Z39_89final.pdf>

Network Development and MARC Standards Office (n.d.). National Level Record---Bibliographic Full Level & Minimal Level. Washington, DC: Library of Congress. Available URL: <http://www.loc.gov/marc/bibliographic/nlr/>

Program for Cooperative Cataloging. (2003). Chart of PCC BIBCO Core Record Standards. Available URL: <http://www.loc.gov/catdir/pcc/bibco/core2002.html >

Tennant, R. 1998. Interoperability: The Holy Grail. Library Journal. Available URL: <http://www.libraryjournal.com/article/CA156495>

Graphics for Paper

An Extensible Approach to Interoperability Testing: The Use of Special Diagnostic Records in the Context of Z39.50 and Online Library Catalogs by William E. Moen, Ph.D., et al.

Table 1. Z-Interop Testbed Search Types

Level 0 Searches	Level 1 Searches
Author Search – Keyword	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	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	Subject Search – Keyword with Right Truncation
	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	Any Search – Keyword with Right Truncation

Table 2. Types of Materials Described by MARC Bibliographic Records

Leader/06 Code	Semantics	AACR Categories of Materials
a	Language material	Books, Pamphlets, and Printed Sheets
c	Notated music	Music (Notated and manuscript music)
d	Manuscript notated music	Music (Notated and manuscript music)
e	Cartographic material	Cartographic Materials
f	Manuscript cartographic material	Cartographic Materials
g	Projected medium	Motion pictures and video-recordings (including digital and non-digital)
i	Nonmusical sound recording	Sound Recordings (musical and non-musical)
j	Musical sound recording	Sound Recordings (musical and non-musical)
k	Two-dimensional nonprojectable graphic	
m	Computer file	Electronic Resources
o	Kit	
p	Mixed material	Graphic materials (includes mixed materials, with or without archival control)
r	Three-dimensional artifact or naturally occurring object	Three Dimensional Artifacts and Realia

Leader/06 Code	Semantics	AACR Categories of Materials
t	Manuscript language material	Manuscripts (including manuscript collections)
Leader/07 Code	**Semantics**	**AACR Categories of Materials**
s	Serial	Continuing Resources

Table 3. Top 5 MARC Indexable Subfields

# of Occurrences	Marc 21 Field	Subfield	Description	Index
602,362	650	a	Subject added entry Topical Term Subfield a = Topical term or geographic name as entry element	Subject
419,641	245	a	Title Statement Subfield a = Title	Title
329,796	245	c	Title Statement Subfield c = statement of responsibility	Author
326,867	650	x	Subject added entry Topical Term Subfield x = General subdivision	Subject
318,692	100	a	Main entry Personal Name Subfield a = personal name	Author

Figure 1. Interoperability Testing Framework

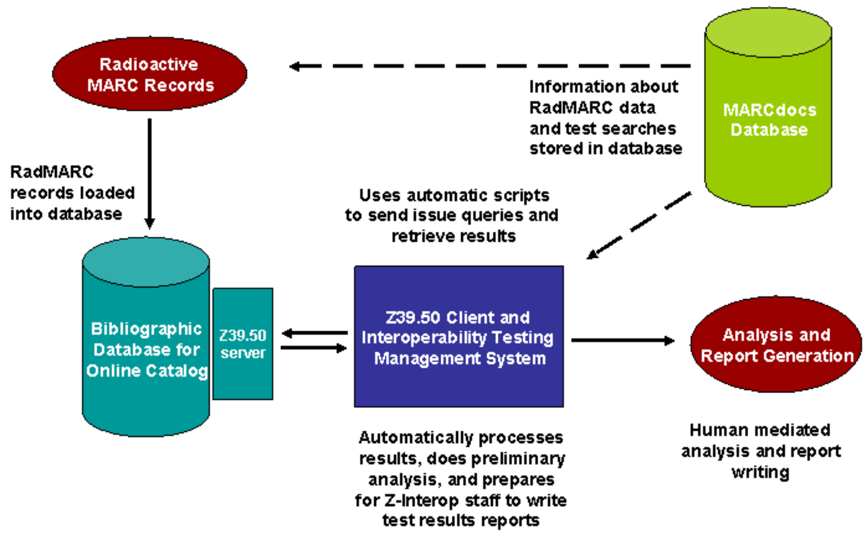

Figure 2. Sample RadMARC Record

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 Dr. William E. Moen at the Texas Center for Digital Knowledge, University of North Texas. Contact Dr. Moen via email <wemoen@unt.edu> for information about this project. Funding for this project is provided by a National Leadership Grant from 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 MARC record. The record support test searches for author, title, subject, and any Bath and U.S. National Z39.50 profile-defined searches, Levels 0 and 1, 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 discovered in a separate analysis. 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.

Figure 3. MARCdocs Database Interface