

Radioactive MARC Records Specifications

Draft – January 1, 2005
[Supersedes December 7, 2004 Draft]

Prepared by
William E. Moen

Contents

1. Introduction	1
2. Types of Records	2
3. Record Identification and Version Information	3
3.1 Use of the 001	3
3.2 Use of the 040	3
3.3 Use of the 583	4
3.4 Example of Use	5
4. Format of the Tokens	5
5. RadMARC Records Required to Test Author, Title, Subject, and Any Searches.....	7
6. RadMARC Records Required to Test Standard Identifier Number Searches	7
7. RadMARC Records Required to Test Qualifying Searches	7
8. Record Creation	7
9. Testing and Editing the RadMARC Records in the Z-Interop Server	8
10. Creating a File of RadMARC Records for Other Targets	8
Appendix A: List of Content Designation for Threshold of Occurrence Code 1.....	9
Appendix B: List of Content Designation for Threshold of Occurrence Code 2.....	10
Appendix C: List of Content Designation for PCC BIBCO Records.....	11

1. Introduction

For the Z-Interop2 Project, a set of special diagnostic MARC records (radioactive or RadMARC records) for conducting interoperability testing are required. The document, *Z39.50 Interoperability Testing Framework for Online Library Catalogs Using Radioactive MARC Records* describes an overview of the interoperability testing approach being used in Z-Interop2. Other preliminary documents have identified search requirements and token requirements for the data values of MARC fields and subfields.

Two Z39.50 profiles listed below provide the specifications that are the basis for the test searches used in the interoperability testing:

- ANSO/NISO Z39.89, The U.S. National Z39.50 Profile for Library Applications, <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) <<http://www.collectionscanada.ca/bath/tp-bath2-e.htm>>

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 should be included in the record.

This document provides the preliminary specifications for the different RadMARC records to be created for use in the Z-Interop2 interoperability testbed. Experience with these records may result in revisions to the specifications.

2. Types of Records

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. The table below shows 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.

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
t	Manuscript language material	Manuscripts (including manuscript collections)
Leader/07 Code	Semantics	AACR Categories of Materials
a	Monographic component part	
b	Serial component part	
c	Collection	
d	Subunit	
i	Integrating resource	
m	Monograph/item	
s	Serial	Continuing Resources

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 proposal for extension to the Z-Interop Project, we proposed to deal with the formats of material as addressed by AACR. These are found 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 will be created will address those in the table below, which also provides the values for the Leader/06 and Leader/07 as appropriate.

Category of Material Described by RadMARC record	Leader/06 Value	Leader/07 Value
Books, Pamphlets, and Printed Sheets	a	
Continuing Resources	a	s
Music (Notated and manuscript music)	c	
Cartographic Materials	e	
Motion pictures and video-recordings (including digital and non-digital)	g	
Sound Recordings (musical and non-musical)	j	
Electronic Resources	m	
Graphic materials (includes mixed materials, with or without archival control)	p	
Three Dimensional Artifacts and Realia	r	
Manuscripts (including manuscript collections)	t	

While there will likely be commonality of many fields across the types of records, and in the indexable fields to support the profile-defined searches, we will create at least one instance of a RadMARC record for each of the 10 types of records.

3. Record Identification and Version Information

Each RadMARC record will be uniquely identified and each version of the record will be noted. The following methods for identifying and describing each record within the record will be used.

3.1 Use of the 001

MARC 001 is the Control Number field, a non-repeatable field with the following definition:

The control number assigned by the organization creating, using, or distributing the record. The MARC code for the organization is contained in field 003 (Control Number Identifier)

This is a standard field in MARC records. In some systems, when a record from some other cataloging source is loaded into the bibliographic database, the number in the 001 is moved to the 035. The 035 is the System Control Number field, a repeatable field with the following definition:

A control number of a system other than the one whose control number is contained in field 001 (Control Number), field 010 (Library of Congress Control Number) or field 016 (National Bibliographic Agency Control Number).

Not all systems move the number from the 001, especially if it is an OCLC number. (Once we get the RadMARC records created, we will contact OCLC to see if they would be interested in allowing us to put these records into the WorldCat database, at which time they would receive a true OCLC number.)

For the initial set of RadMARC records, the Control Number will have a coded structure as follows:

- Characters 1-10: Project signature for the record; value = UNTRadMARC
- Characters 11-13: Sequential number with left padding; value begins at 001 and continues sequentially for subsequent records.

An example of this is: UNTRadMARC001, UNTRadMARC002, etc.

3.2 Use of the 040

MARC 040 is the Cataloging Source field, a non-repeatable field with the following definition:

The MARC code for or the name of the organization(s) that created the original bibliographic record, assigned MARC content designation and transcribed the record into machine-readable form, or

modified (except for the addition of holdings symbols) an existing MARC record. These data and the code in 008/39 (Cataloging source) specify the parties responsible for the bibliographic record.

MARC 040 defines several subfields, and the RadMARC records will use only subfield a, defined as Original Cataloging Agency. For the initial set of RadMARC records, the Cataloging Source will have a code to represent the Z39.50 Interoperability Testbed Project in the form of:

- ZinteropUNT

When loaded into a remote system, this value is not likely to be changed and provides one more place to indicate the source of these special MARC records.

3.3 Use of the 583

To record version and other information about each RADMARC records, we propose to use MARC 583, the Action Note field, a repeatable field with the following definition:

A copy-specific field that contains information about processing, reference, and preservation actions.

MARC 583 has a sufficient number of subfield structures to enable the recording of the following types of information:

- Signature
- Identifier
- Version
- Description
- Scheme
- Threshold
- Record Creator

This, however, means that we will revise the meaning of the 583 subfield codes as follows:

- a - Action
 - **RadMARC semantics: <signature>** a constant string -- that is always the same for every RadMARC record.
 - Value = RadMARC
- b - Action identification
 - **RadMARC semantics: <identifier>** any string that is "as unique as possible". Will use domain name for project <www.unt.edu/zinterop> and a sequential number (which is the same numeric value as in the 001).
 - Value = www.unt.edu/zinterop/001, www.unt.edu/zinterop/002, etc.
- d - Action interval
 - **RadMARC semantics: <version>** is an integer, beginning at 1 for the initial version of the record and incremented by 1 in each subsequent version.
- e - Contingency for action
 - **RadMARC semantics: <scheme>** identifies uniquely the type of search or other parameters describing the record in the context of interoperability testing.
 - Value will be defined for the RadMARC records produced by the Z-Interop2 Project.
- i - Method of action
 - **RadMARC semantics: <scheme dependent information>** includes scheme dependent information regarding this particular record.
 - Value will be defined for the RadMARC records produced by the Z-Interop2 Project.
- \$k - Action agent (R)
 - **RadMARC semantics: <record creator>** contains the name of the person creating the RadMARC record
 - Value = name of person creating record

- x - Nonpublic note
 - **RadMARC semantics: <description>** is arbitrary human-readable text describing the provenance and intent of the record, and perhaps even including things like a version-control log describing its history.
 - Value will be standardized to the extent possible along the lines of:
"This is a specially created test record 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>. 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 1 for occurrences of more than 100,000 times, and this is the first version of this record."

3.4 Example of Use

The following shows an example of a RadMARC record using the above specifications for identification, version, and other information:

```
001 UNTRadMARC001
040 $a ZinteropUNT
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 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>. 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.
```

4. Format of the Tokens

The core of the RadMARC records is specially designed tokens that populate each indexable field/subfield. 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 RadMARC field/subfield. A field/subfield may have a sequence of tokens. 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 to represent the 10 categories of materials discussed in Section 2 as follows:
 - Leader/06 **a** - Language material
Includes printed, microform, and electronic language material.
 - Leader/06 **c** - Notated music
Includes microform and electronic notated music.
 - Leader/06 **e** - Cartographic material
Includes maps, atlases, globes, digital maps, and other cartographic items.
 - Leader/06 **g** - Projected medium
Examples include: motion pictures, video recordings (including digital video), filmstrips, slides, transparencies, or material specifically designed for projection.
 - Leader/06 **j** - Musical sound recording
Examples include: phonodiscs, compact discs, or cassette tapes.
 - Leader/06 **m** - Computer file
Includes the following classes of electronic resources: computer software (including programs, games, fonts), numeric data, computer-oriented multimedia,

online systems or services. For these classes of materials, if there is a significant aspect that causes it to fall into another Leader/06 category, the code for that significant aspect is used instead of code m (e.g., vector data that is cartographic is not coded as numeric but cartographic). Other classes of electronic resources are coded for their most significant aspect (e.g., language material, graphic, cartographic material, sound, music, moving image). In case of doubt or if the most significant aspect cannot be determined, consider the item a computer file.

- Leader/06 **p** - Mixed material
Indicates that there are significant materials in two or more forms that are usually related by virtue of their having been accumulated by or about a person or body. Includes archival fonds and manuscript collections of mixed forms of materials, such as text, photographs, and sound recordings.
- Leader/06 **r** - Three-dimensional artifact or naturally occurring object
Includes man-made objects, such as models, dioramas, games, puzzles, simulations, sculptures and other three-dimensional art works and their reproductions, exhibits, machines, clothing, toys, and stitchery, and naturally occurring objects, such as microscope specimens and other specimens mounted for viewing.
- Leader/06 **t** - Manuscript language material
- Leader/07 **s** - Indicating a serial or continuing resource
- 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
 - Note: Even though some subfields in specific fields can repeat, the current token format does not allow specification of the occurrence number of the subfield within the field.
- 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 Monograph-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

If there was a second instance of this field (in this example the 245), the token element for the second occurrence would be `rm2452a11r`. An example of a complete MARC 245 with a sequence of tokens in selected subfields follows:

```
245 $a rm2451a1r rm2451a2r rm2451a3r $b rm2451b1r rm2451b2r rm2451b3r $c
rm2451c1r rm2451c2r
```

In any given subfield, a sequence of tokens will not exceed three tokens. The limitation allows the appropriate interoperability testing of various phrase-oriented searches as defined in the profiles without adding undue complexity to the RadMARC records.

5. RadMARC Records Required to Test Author, Title, Subject, and Any Searches

A number of sets of RadMARC records are required that will support the testing of profiled-defined author, title, subject, and any searches at Levels 0 and 1. In each set, there will be 10 separate records reflecting the different types of records or materials described. Each record in each set will contain the exact same tokens except for the second character of the token element that will reflect the type of record or material. For example, the token element in a 245 field for:

- A Books, Pamphlets, and Printed Sheets Language Material type record will be:
 - ra2451a11r ra2451a21r ra2451a31r (which uses the Leader/06 code “a”)
- A Music (Notated and manuscript music) type record will be:
 - rc2451a11r rc2451a21r rc2451a31r (which used the Leader/06 code “c”)
- A Continuing Resource type record will be:
 - rs2451a11r rs2451a21r rs2451a31r (which used the Leader/07 code “s”)

Our goal is to create several sets of RadMARC records that will have increasing levels of content designation (i.e., the number of fields and subfields) included. The level of content designation for each set of records will be based on the following specifications (which may evolve based on experience):

- The threshold of occurrence of indexable fields that resulted from previous analysis undertaken in the first phase of the Z-Interop Project. The threshold of occurrence will determine which content designation gets populated in each set of records. The threshold of occurrence sets will be defined as follows as indicated by Threshold of Occurrence Code held in the 583 \$i. :
 - 1 = The 19 most commonly occurring indexable fields for author, title, and subject-related data as identified in the previous research. See Appendix A for a list of these fields/subfields.
 - 2 =The indexable fields for author, title, and subject-related data occurring at least 1,000 times as identified in the previous research; approximate number of these fields/subfields is: **XXX**. See Appendix B for a list of these fields/subfields.
 - 3 =All indexable fields for author, title, and subject-related data as identified in Z-Interop Project indexing guidelines; approximate number of these fields/subfields is: **XXX**. See the Z-Interop Indexing Guidelines for the complete list of these fields/subfields.
- The indexable fields as listed in the Z-Interop Indexing Guidelines that are recommended by the Program for Cooperative Cataloging Bibliographic Core (BIBCO) Record Elements; approximate number of these fields/subfields is: **XXX**. See Appendix C for a list of these fields/subfields.

With this arrangement, there will be four sets of RadMARC records, with each set containing 10 records.

6. RadMARC Records Required to Test Standard Identifier Number Searches

[to be completed]

7. RadMARC Records Required to Test Qualifying Searches

[to be completed]

8. Record Creation

Project staff will create RadMARC records according to the above specifications using MARC creation software that produces reliable and valid MARC records on export. Initially, the software to be used is MARC Magician, a PC-based application that provides the functionality needed.

Priority of creation should be on the sets of records at Threshold of Occurrence Code 1 and the BIBCO specifications. The time needed to create each set of records should be recorded to get an understanding of the level of effort needed.

9. Testing and Editing the RadMARC Records in the Z-Interop Server

The RadMARC records will be loaded into the Z-Interop Testbed Server. The RadMARC records can be checked by searching the Z-Interop server with the Bookwhere client. The Bookwhere client allows us to set the attribute combination for each of the profile-defined searches.

MARC Magician, unfortunately, places the phrase [Demo Record] in each subfield that is populated. This phrase will need to be edited out of each RadMARC record. The Sirsi Workflows client can be used to easily edit the RadMARC record. Once the records have been edited, the Index Data test harness can be used to run a trial set of searches and generate reports.

10. Creating a File of RadMARC Records for Other Targets

Once the testing has been done on the sets of RadMARC records, and any issues or problems with the records have been resolved or edited away, each set of RadMARC records will be exported from the Z-Interop testbed server to create the authoritative file of RadMARC records that can be used by other targets to load into their bibliographic database for interoperability testing.

Appendix A: List of Content Designation for Threshold of Occurrence Code 1

Based on previous analysis, the Z-Interop Project, Phase 1, identified the following 19 fields/subfields from the Z-Interop Indexing Guidelines as the most commonly occurring. This set of content designation will be used to create the most minimal versions of the RadMARC records.

# 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
231,459	650	z	Subject added entry Topical Term Subfield z = Geographic subdivision	Subject
176,916	700	a	Added entry Personal Name Subfield a = personal name	Author
169,178	245	b	Title Statement Subfield b = Remainder of title	Title
149,540	100	d	Main entry Personal Name Subfield d = dates associated with a name	Author
118,647	651	x	Subject added entry Geographic Name Subfield x = General subdivision	Subject
113,050	651	a	Subject added entry Geographic Name Subfield a = Geographic name	Subject
83,607	650	v	Subject added entry Topical Term Subfield v = Form subdivision	Subject
74,606	700	d	Added entry Personal Name Subfield d = dates associated with a name	Author
69,636	600	a	Subject added entry Personal Name Subfield a = personal name	Subject
66,375	710	a	Added entry Corporate Name Subfield a = corporate name or jurisdiction name	Author
64,433	440	a	Series Statement Added Entry Title Subfield a = title	Title
62,853	490	a	Series Statement Subfield a = Series statement	Title
56,229	600	d	Subject added entry Personal Name Subfield d = dates associated with a name	Subject
55,311	653	a	Index Term Uncontrolled Subfield a = the term	Subject

Appendix B: List of Content Designation for Threshold of Occurrence Code 2

[to be completed]

Appendix C: List of Content Designation for PCC BIBCO Records

[to be completed]