

# **MULTIPLE LANGUAGES, MULTIPLE SCRIPTS, AND MULTIPLE HEADINGS IN MARC AUTHORITY RECORDS**

**prepared by**

**Randall K. Barry  
(Internet: [RBAR@LOC.GOV](mailto:RBAR@LOC.GOV))**

**Library of Congress  
Acquisitions and Bibliographic Access Directorate**



## MULTIPLE LANGUAGES IN MARC

- **It is common for MARC records to have data in more than one language;**
- **Transcribed descriptive data is generally in the *language(s) of the item*;**
- **Access points and notes are generally in the *language of the catalog*;**
- **Some libraries provide access points and notes in more than one language (examples: Canada, Switzerland)**

## LANGUAGE IDENTIFICATION IN MARC

- **Language of the item (in bibliographic records);**
  - 1 Code in field 008/35-37 and others in field 041;
  - Textual note in field 546;
- **Language of the heading (in MARC authority format 008/35-37) was made obsolete;**
- **Language of catalog is coded in field 040, subfield \$b;**
- **No explicit field-level language flag is currently defined for MARC 21 authority or bibliographic records;**

## LANGUAGE CONTEXT IN MARC AUTHORITIES

- **Language of the heading has been treated as an *implicit characteristic*, and is not coded *explicitly* in MARC;**
- **The issue of an explicit context (language/nationality) code has been discussed (field-level?, record level?);**
- **In many countries, coding accurately for language and/or nationality would be very difficult;**
- **How would you code persons and entities active in several countries, and/or international agencies?**

## MULTILINGUAL MARC AUTHORITY RECORDS

- **Most catalogs having headings in many languages;**
- **Many name authority records have tracings from different language forms;**
- **These include names of famous people in translation;**
- **Most international bodies;**
- **Government bodies with international scope (national foreign affairs ministries, embassies, military units);**
- **Catalogs of bilingual and multilingual countries.**

## LANGUAGE ASSUMPTIONS IN MARC

- **The language of the catalog is not necessarily the same as the language of the items represented in it;**
- **Transcribed data is usually in the original language;**
- **Translation into the language of the catalog is provided for some data (certain names, titles, etc.);**
- **National and local cataloging rules govern the language of the catalog;**
- **Thesauri used are usually in the language of the catalog**

## LIMITATIONS OF EARLY MARC CATALOGS

- **Computer hardware and software in the 20th century did not support all languages equally;**
- **Many computer systems had an English language bias;**
- **Input & retrieval interfaces were often English only;**
- **Extensions to the basic Latin (Roman) alphabet were limited;**
- **Support for non-Latin (non-Roman) scripts was limited**

## EARLY SOLUTIONS

- **Fully transliterated (romanized) cataloging records;**
- **Preference given to non-vernacular (not original) forms of name and title;**
- **At best, tracings and references made from others;**
- **Special techniques were used to encode unusual, modified Latin script letters (using double underscore);**
- **Some manual catalogs were not retrospectively converted to MARC at all.**



## MULTIPLE SCRIPTS IN LIBRARIES

- **Historically, many manual library catalogs included information in a variety of scripts:**
  - Latin (Roman) script was very common;
  - Greek and Hebrew script was also common;
  - Arabic, Chinese, Indic, and other modern scripts could be found in larger libraries;
  - Lesser-known scripts were generally limited to their countries of origin and special libraries elsewhere.

## MULTIPLE SCRIPTS IN MARC

- **Development of the MARC formats began with Latin (Roman) script data; non-Latin data was omitted;**
- **The need to address non-Latin cataloging data was recognized early on;**
- **In 1979 a decision was made to fully transliterate non-Latin cataloging for certain languages;**
- **Japanese, Arabic, Chinese, Korean, Persian, Hebrew, & Yiddish (J-A-C-K-P-H-Y) were excluded.**

## JACKPHY AND MARC CHARACTER SETS

- **Special MARC character sets were developed to support the non-Latin "JACKPHY" scripts/languages;**
- **"Vernacular" MARC cataloging for JACKPHY began in 1982;**
- **The Research Libraries Information Network (RLIN) system developed a JACKPHY input/update system;**
- **MARC record creation was based on the "dual script" concept.**

## EARLY MARC 8-BIT CHARACTER SETS

- **In the old "MARC-8" (8-bit) environment, only eight modern writing systems were accommodated:**
  - Latin (Roman) script (for many modern languages)
  - Greek script (for modern and classical Greek)
  - Hebrew script (for Hebrew and Yiddish)
  - Arabic script (for Arabic, Persian, Urdu, etc.)
  - Cyrillic script (for Russian, Serbian, Ukrainian, etc.)
  - East Asian scripts (for Chinese, Japanese, and Korean)

## MARC AND UNICODE™

- **The MARC-8 character sets were fully mapped to Unicode in the late 1990s;**
- **Implementation of Unicode required coordination between many large MARC record producers;**
- **Most major MARC systems implemented Unicode sometime after January 2000;**
- **OCLC and RLIN implemented Unicode by 2004;**
- **The Library of Congress implemented Unicode in 2005.**

## CURRENT MARC CHARACTER LIMITATIONS

- **MARC 21 users have been asked to limit their use of Unicode to the characters that map to MARC-8;**
- **In practice, many libraries have already expanded their use to many new scripts (e.g., Armenian, Thai, Indic);**
- **This "de facto" expansion is spreading quickly;**
- **The MARC community is discussing options for dealing with libraries using all of Unicode;**
- **Most MARC 21 systems allow full Unicode use.**

## DUAL SCRIPT MARC RECORDS

- **Non-Latin script data is transliterated into the Latin script and transcribed in the normal MARC fields;**
- **This approach was compatible with the solution for non-JACKPHY scripts (Greek, Cyrillic, etc.);**
- **"Vernacular" script data would be repeated in separate occurrences of the MARC fields;**
- **Repeated data would be embedded in field 880 for ease of suppression in display and printing.**

## FIELD 880 EMBEDDING TECHNIQUE

100 1[] \$6880-01\$aTheodōropoulou, Viky,\$d1958-

880 1[] \$6100-01/(S\$aΘεοδωροπούλου, Βίκυ,\$d1958-

↑ Field tag

↑ Indicators required by corresponding data

↑ Control subfield \$6

↑ Variable field data

- Subfield \$6 includes the tag of the embedded field, a sequencing number, and a character set identifier.



## **MODEL A: TRANSLITERATION & VERNACULAR**

- **Cataloging data in the script of the catalog is recorded in the normal (non-embedded) MARC fields;**
- **Representation of the same data in an alternative (usually the original) script is embedded in field 880;**
- **Field 880 is repeated for each embedded field needed;**
- **Many systems cannot process data in field 880 easily;**
- **Currently, field 880 is used mostly for non-Latin data.**

## EXAMPLE OF A MODEL A RECORD

**040** **[][]\$aDLC\$cDLC**

**066** **[][]\$c(S**

**100** **1[]\$6880-01\$aBekatōros, Stephanos,\$d1946-**

**400** **1[]\$6880-00\$aVekatōros, Stefanos,\$d1946**

**880** **1[]\$6100-01/(S\$aΜπεκατώρος, Στέφανος,\$d1946-**

**NOTE:** A field can be linked to nothing, in which case the sequence number in subfield \$6 is "-00".

## ADVANTAGES OF MODEL A

- **Easier suppression of non-Latin data in Latin script cataloging environments;**
- **Clear distinctions are made between the data in different scripts (vernacular & non-vernacular);**
- **It is consistent with the cataloging practice for fully transliterated scripts in Latin script catalogs;**
- **This practice avoided the issue of repeating non-repeatable fields, such as field 100 or field 245.**

## **DISADVANTAGES OF MODEL A**

- **Marginalizes the alternative graphic representation of cataloging data, which is usually the original script;**
- **Field 880 embedding technique in MARC is complex;**
- **Linkages between non-embedded and embedded data in field 880 are difficult to input manually;**
- **Implementation has been limited mainly to U.S. systems (RLIN, OCLC) and never embraced by the international library (MARC user) community.**

## **MODEL B: SIMPLE MULTISCRIPPT RECORDS**

- **All data is contained in the regular MARC variable data fields;**
- **Whatever script is needed is used wherever needed;**
- **No special embedding technique is used;**
- **No special linkages between data in different scripts is required;**
- **No character set identification is made at the field level.**

## EXAMPLE OF MODEL B AUTHORITY RECORD

**040** [[]]\$aRuMoRGB\$brus\$сRuMoRGB

**100** 1[]\$aВасниев, К. С.\$q(Каплан Сафербиевич)

**400** 1[]\$aВасниев, Каплан Сафербиевич

**400** 1[]\$aVasniev, K. S.\$q(Kaplan Saferbievich)

**400** 1[]\$aVasniev, Kaplan Saferbievich

**670** [[]]\$aКто есть кто в нефт. и газ. пром., 1992:  
\$b(К.С. Васниев) колофон (Каплан Сафербиер-вич  
Васниев; Kaplan Saferbievich Vasniev)

## ADVANTAGES OF MODEL B

- **The original (often non-Latin) script is treated equally with the alternative (usually Latin) script;**
- **No special field embedding must be contended with;**
- **Scripts can be mixed much more effectively at the field and subfield level;**
- **Since there are no linkages between fields, record creation and maintenance is much easier.**

## **DISADVANTAGES OF MODEL B**

- **Identification and suppression of data in specific scripts is more difficult;**
- **Relationship between data in different scripts is not identified by explicit linkages (in subfield \$6);**
- **The script is not identified at the record or field level;**
- **It is simpler to mix scripts unwittingly within strings, sometimes using characters from one script with those of another (Greek "H", Latin "H", and Cyrillic "H")**



## **CURRENT TRENDS IN MARC IMPLEMENTATIONS**

- **Anglo-American libraries are continuing to create dual script records following Model A for bibliographic data only, and only for the JACKPHY scripts;**
- **Non-English cataloging agencies overwhelmingly prefer Model B for handling non-Latin scripts;**
- **A switch to Model B has been discussed and is likely to occur in Anglo-American libraries for non-JACKPHY;**
- **Model A is being considered for non-Latin authority data, perhaps as early as 2007 for all scripts.**

## **RATIONALE FOR PREFERRING MODEL B**

- **Countries that use scripts other than Latin prefer the simple multiscript record approach;**
- **The implementation of Unicode worldwide has solved the problem of non-Latin script support in libraries;**
- **The use of MARC 21 field 880 is complex;**
- **Only a relatively small number of libraries in the U.S. currently use field 880;**
- **The Model B approach will be key to international cooperation in the area of authority control.**

## **MULTIPLE HEADINGS IN MARC AUTHORITIES**

- **Most library catalogs have one authoritative form for each heading;**
- **Catalogs in some countries require multiple authoritative forms (bilingual/multilingual countries);**
- **Multiple authoritative forms are usually handled by multiple MARC authority records;**
- **Multiple MARC authority records are linked by See-Also Tracings or Heading Linking Entries (7XX).**

## **SEE-ALSO TRACINGS TO MULTIPLE HEADINGS**

- **It is always possible to link between two valid authoritative forms of name with see-also tracings;**
- **Relationships between earlier and later forms of name are usually handled this way;**
- **Complex see and see-also references can also be used to relate headings;**
- **Tracings and references are generally used in monolingual catalogs.**

## EXAMPLE OF SEE-ALSO TRACING METHOD

**010** [ ] [ ] \$an 78004438

**110** 2 [ ] \$aInternational Federation of Library Associations  
and Institutions

**410** 2 [ ] \$aIFLA

**410** 2 [ ] \$aDiethnēs Enōsē Vivliothēkonomikōn Enōseōn  
kai Organismōn

**410** 2 [ ] \$aΔιεθνής Ένωση Βιβλιοθηκονομικών Ενώσεων  
και Οργανισμών

**510** 2 [ ] \$wa\$aInternational Federation of Library  
Associations *[see-also tracing from earlier name]*

## EXAMPLE OF SEE-ALSO TRACING METHOD: CORRESPONDING MARC AUTHORITY RECORD

**010** [ ] [ ] \$an 81032757

**110** 2 [ ] \$International Federation of Library Associations

**410** 2 [ ] \$aIFLA

**510** 2 [ ] \$wa\$aInternational Federation of Library  
Associations and Institutions *[see-also tracing from later  
name]*

## HEADING LINKING ENTRY TECHNIQUE

- Provides heading linkages to other authoritative forms;
- Other authoritative forms may be in separate databases (catalogs) or thesauri;
- Links may be to forms in a different language/script;
- Links may be to headings formulated according to different cataloging rules;
- Heading linking entries may duplicate forms also given as tracings or references in an authority record.

## EXAMPLE OF A HEADING LINKING ENTRY

**016 [][]\$a0005G9610E**

**100 0[]\$aJohn Paul\$bII,\$cPope,\$d1920-2005**

**400 1[]\$aWojtyła, Karol Józef,\$d1920-2005**

**400 0[]\$aJoannes Paulus\$bII,\$cPope,\$d1920-2005**

**510 2[]\$aCatholic Church.\$bPope (1978-2005 :**

**John Paul II)\$0(CaOONL)0005G9580E**

**700 05\$aJean Paul\$bII,\$cpape,\$d1920-2005**

**\$0(CaOONL)0005G9610F**

*[heading linking entry to the French language form]*



**EXAMPLE OF A HEADING LINKING ENTRY:  
CORRESPONDING MARC AUTHORITY RECORD**

**016 [][]\$a0005G9610F**

**100 0[]\$aJean Paul\$bII,\$cpape,\$d1920-2005**

**400 1[]\$aWojtyła, Karol Józef,\$d1920-2005**

**400 0[]\$aJoannes Paulus\$bII,\$cpape,\$d1920-2005**

**510 2[]\$aÉglise catholique.\$bPape (1978-2005 :**

**Jean Paul II)\$0(CaOONL)0005G9580F**

**700 05\$aJohn Paul\$bII,\$cPope,\$d1920-2005**

**\$0(CaOONL)0005G9610E**

*[heading linking entry to the English language form]*

## LIMITATIONS OF HEADING LINKING ENTRIES

- **7XX fields contain data that can be used to provide access to alternative authoritative forms of name;**
- **7XX fields do not specify how a system handles the headings they contain;**
- **Functionality of the 7XX fields is dependent on the sophistication of a MARC system's design;**
- **Behavior of catalog users must be taken into account;**
- **Retraining of catalog users may be needed.**

## USE OF 7XX FIELD IN THE NORTH AMERICA

008 060417nb[]acannaaan[]ua[]ana[]

010 [][]\$an 2006269255

040 [][]\$aDLC\$beng\$cdLc\$dHkCAN [*subfield \$f not needed*]

110 2[]\$aBeijing da xue.\$bTu shu guan

410 2[]\$a北京大学.\$b图书馆

410 2[]\$aΠανεπιστήμιο Πεκίνου.\$bΒιβλιοθήκη

410 2[]\$aPeking University.\$bLibrary

710 27\$a北京大学.\$b图书馆\$0(HkCAN)0012345 \$2hkcan

**Note: 008/11 (Subject heading system/thesaurus) = "a" (LCSH)**

## USE OF 7XX FIELD IN CHINA

008 060417nb[]acznnaaan[]ua[]ana[]

035 [][]\$a(HkCAN)012345

040 [][]\$aDLC\$bchi\$cDLC\$dHkCAN\$fhkcan [*thesaurus code*]

110 2[]\$a北京大学.\$b图书馆

410 2[]\$aBeijing da xue.\$bTu shu guan

410 2[]\$aΠανεπιστήμιο Πεκίνου.\$bΒιβλιοθήκη

410 2[]\$aPeking University.\$bLibrary

710 20\$aBeijing da xue.\$bTu shu guan\$0(DLC)n[]2006269255

**Note: 008/11 (Subject heading system/thesaurus) = "z" (Other);  
thesaurus identified in 040\$f**