

Frequently Asked Questions : Cataloging of Nonroman Script Materials at the Library of Congress in a Post-RLIN World

Q1. Where will LC's nonroman catalog records be created and maintained after RLIN 21 goes away?

A1. LC will migrate the cataloging of monographs from RLIN 21 to LC's local catalog (the Voyager system from Ex Libris) in late August, 2007. LC's nonroman catalogers will also be using an "add on" tool compatible with Voyager developed by Gary Strawn (Northwestern University) and David Bucknum (Library of Congress) that provides a cataloger-friendly interface for pairing regular MARC fields with their 880 field counterparts, as well as providing transliterations between roman and nonroman forms for some languages and scripts. Nonroman serials and integrating resources will continue to be cataloged as part of the CONSER program on OCLC.

Q2. What languages or scripts will LC catalogers use?

A2. For this transition period, records for monographs cataloged with nonroman scripts will continue to be the same as in RLIN 21: Chinese, Japanese, Korean, Hebrew, Arabic, Persian and Yiddish. Serial bibliographic records include these languages, as well as Cyrillic and Greek, i.e., the full MARC-8 Repertoire. Management decisions regarding the expansion of LC monographic cataloging to include Cyrillic and Greek scripts are pending at this time. LC is also studying the question of supplying nonroman data for those scripts outside of the MARC-8 Repertoire, but no decisions on expansion are available at this time.

Q3. When an LC cataloger decides to provide nonroman scripts, which MARC fields are provided in both romanized and nonroman paired fields?

A3. To provide a complete catalog record in the roman alphabet, LC always provides roman alphabet forms of most fields. Under certain circumstances, LC also provides parallel nonroman forms. Although the exact practices may vary among language groups as indicated, in general the following guidelines are applicable for full level cataloging:

Fixed fields : no nonroman data

010-082, 255, 263, 300 : no nonroman data

1XX, 240 : supply nonroman fields according to language-specific guidelines

245, 250, 260, 4XX : always supply nonroman field if nonroman data are present

5XX : nonroman data generally not supplied, except for Chinese

600, 610, 611, 630, 651 : nonroman fields supplied according to language-specific guidelines

650 : no nonroman data

700, 710, 711, 730, 740 : nonroman fields supplied according to language-specific guidelines

Core level cataloging guidelines for providing paired nonroman script fields provide for fewer paired fields than full level cataloging, see Descriptive Cataloging Manual (DCM) B.16.5.4 for details. See also the PCC Core Level documentation on "Guidelines for Multiple Character Sets" at <http://www.loc.gov/catdir/pcc/bibco/coreintro.html#9> for more information. After some experience cataloging nonroman script materials in LC's

local catalog, and further experience with automatic transliteration software, LC may reevaluate its practices with respect to the number of nonroman fields added.

Q4. For languages and scripts that are read from right-to-left, e.g., Hebrew, Arabic, Persian, and Yiddish, which Unicode Formatting Characters are used in LC records?

A4. LC's practices for Unicode Formatting Characters (UFCs) in UTF-8 bibliographic records containing right-to-left scripts are generally the same as applied in RLIN 21:
--Supply Right-to-Left marks (RLM, U+200F) before each subfield delimiter (except \$6) and after each subfield code (except \$6) in 880 fields with right-to-left orientation
--When necessary to embed left-to-right data in fields oriented as right-to-left (e.g., dates in 260\$c in an 880 field containing Arabic or Hebrew script), catalogers will add a Left-to-Right Embedding character (LRE, U+202A) before entering the left-to-right data string, and a Pop Directional Formatting character (PDF, U+202C) after the data string. LC has no plans at this time to use additional UFCs.

For subscribers who receive monographic nonroman records from the Cataloging Distribution Service (BOOKS-ALL), records are available in MARC-8 or UTF-8. For MARC-8 records all UFCs are eliminated from the records; for UTF-8, UFCs will be present only in records created or updated once LC begins working in its own catalog in August of 2007.

Q5. Are there any plans to use nonroman data in authority records?

A5. The major authority record exchange partners that serve as nodes for the LC/NACO Authority File have agreed to a basic outline that will allow for the addition of nonroman references in name authority records distributed as part of the NACO program. It is expected that the use of nonroman data in authorities will begin no earlier than April of 2008. Announcements regarding policies and practices will be made prior to that time. The Library of Congress currently has no plans to add nonroman data to subject authority records for the Library of Congress Subject Headings.

Q6. What spacing conventions will be used in Chinese, Japanese, and Korean records in the LC catalog?

A6. In June 2007 LC announced a change in practice in the use of spaces in nonroman 880 fields in bibliographic records for Chinese, Japanese, and Korean. By the beginning of September 2007 LC's Chinese and Japanese records for monographs found in the LC online catalog should follow this new practice, i.e., they will not have spaces between Chinese or Japanese characters. Chinese and Japanese records with the new spacing convention will be re-distributed to MARC Distribution Service customers. More information, including the applicable changes to both monographs and serials, is available at: <http://www.loc.gov/cds/notices/spacingcjk.pdf>

Q7. How can users of the LC online catalog search and display nonroman languages and scripts?

A7. Instructions for setting browsers to enable search and display of nonroman languages and scripts are available at: <http://catalog.loc.gov/help/unicode.htm>

Additional questions can be referred to the Cataloging Policy and Support Office,
cpsol@loc.gov

8/16/2007