Embracing Authenticity – More Scripts, Less Romanization

Committee on Technical Processing Session
CEAL Annual Meeting
Seattle, March 14, 2024
Overview

- History and Current Status
- BF to MARC Conversion
- Provide Transliterated data
- Phased Implementation of BF and FOLIO
- Policy Implications
- Future Recommendations
- Significance of Non-Latin Script Input

LC’s Non-Latin Script Materials

African and Middle Eastern Division holds 650,000 volumes.

Asian Division holds more than 4 million items.

Over one million in Slavic languages in Cyrillic.

Vast holdings in multiple formats (cartographic material, film, music, prints, photos, ephemera, etc.)
Manually Cataloging Non-Latin Scripts (1898-1969)

Early Library Catalogs

Library of Congress distributed printed cards including works in non-Latin scripts.

MARC Standards
Developed by Library of Congress to enable electronic sharing of catalog records.

Limit to Latin Script
LC’s early MARC system limited to Latin script

Non-Latin Languages
Records created from 1969-1981 did not include original scripts
Expansion of Non-Latin Script Cataloging

RLG introduced enhancement for CJK cataloging

RLG brought non-Latin capability to JACKPHY (Japanese, Arabic, Chinese, Korean, Persian, Hebrew, Yiddish)

1983

1991

LC switched non-Latin monograph cataloging for JACKPHY to Voyager.

2007

2023

Armenian and Thai added to the mix.
Non-Latin Script Cataloging at LC (Serials, Authorities)

Serial cataloging

Non-Latin script serial cataloging continued to use OCLC

Name Authorities

In 2008, non-Latin references in name authority file implemented
BIBFRAME and FOLIO

**BIBFRAME**
A data model for bibliographic description, designed to replace the MARC standards

**BF 2.0**
In 2016, LC released version 2.0

**Testing Unicode**
Include scripts beyond JACKPHY+CAT

**Best practices of non-Latin script input, emphasis on original script**

In October 2024, LC will implement initial phase of FOLIO-Acquisitions

In spring 2025, *Inventory* will be implemented for bibliographic description
Best Practices of BIBFRAME Input

Instance
- Descriptive metadata in the original script will be provided
  - Optional transliteration: title proper, other title information, statement of responsibility, and publication statement.
  - Table of contents or summary in original script.

Work
- Descriptive access point, including RDA preferred title will be recorded in transliterated form.
  - Title in original script may be recorded optionally.

Reflect the pre-MARC era approach
BIBFRAME to MARC Conversion

• PCC Guidelines for Providing Bibliographic Description in Multiple Character Sets

• External libraries expect to receive MARC records containing both Latin and non-Latin scripts for their local systems
Providing Automated Transliteration

**ScriptShifter**

An API-powered tool

[TRANSLITERATE!]

https://scriptshifter-test.knowledgetx.com/

**Use of Angle Bracket**

*Missing Transliteration*
Phased Implementation of Non-Latin Script Input in BF and FOLIO

JACKPHY+CAT languages, currently included in Voyager

Phase 1

All the other languages for which transliteration tools may not be available

Phase 2

Languages for which transliteration tools are available in ScriptShifter

Phase 3
Policy
Implications

- Less romanization, emphasizing original scripts whenever feasible
- Use MARC Model B for BIBFRAME input
- Use angle brackets to indicate missing romanization
Future Recommendations

- Expand non-Latin scripts beyond JACKPHY+C in the NACO File
- Integrate Parallelogram functionality into FOLIO
Significance of Effective Non-Latin Script Input

- Promote cultural and linguistic diversity of library collections
- Improve user experience and enhance discoverability
- Meet the needs of users on a global scale
- Promote Unicode international standard for cataloging and metadata
Thank You!

Jessalyn Zoom
Asian and Middle Eastern Division
Library of Congress
jiwu@loc.gov