Policy Issues on Non-Latin Script Input in BIBFRAME of FOLIO

Introduction

This report summaries the issues concerning bibliographic description for non-Latin script materials in the Library of Congress's BIBFRAME Editor. With the imminent implementation of BIBFRAME in FOLIO for non-Latin script materials in spring 2025, there is an urgent need to establish a policy on the extent of transliterated data to be included in BIBFRAME, and subsequently, identify solutions for generating transliterated data to facilitate the BIBFRAME to MARC conversion. The goal of this report is to capture the issues and set the stage for further discussion and actions.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

The PCC Guidelines document was developed in 2009 based on recommendations from the PCC Policy Committee. The bibliographic structure endorsed in the Guidelines uses MARC Model A (Non-Latin script information is recorded in the MARC 880 field with transliteration recorded in a regular MARC field). The document was last revised in 2017. While the decision to include non-Latin script data in PCC bibliographic records is optional, when adding non-Latin script data, these guidelines should be followed.

Current Status of Non-Latin Script Input in the Voyager System

In the current Voyager system, bibliographic records for non-Latin script materials in the JACKPHY+CAT languages (Japanese, Arabic, Chinese, Korean, Persian, Hebrew, Yiddish, Armenian, Thai and most of Cyrillic languages) are created following MARC Model A. Non-Latin script information is recorded in the MARC 880 field with transliteration recorded in the regular MARC field. The 880 field was implemented decades ago to allow non-Latin script data to get around the limitation of library systems on handling data in multiple scripts.

Testing Non-Latin Script Input in BIBFRAME Editor

ABA implemented BIBFRAME Pilot 2 in summer of 2017. Over the past six years, testing for non-Latin script input in the BIBFRAME Editor has been following MARC Model B. Under this model, bibliographic description is created in the original script in the Instance, with access points provided in Latin transliteration in the Work. While the testing has focused on the use of MARC Model B, there has been a lack of clear policy regarding the extent of transliterated data to be included. In anticipation of the implementation of BIBFRAME FOLIO in spring 2025, it is important to establish a policy regarding the amount of transliterated data that should be entered in bibliographic description for non-Latin script materials in BIBFRAME Editor.

Recommendations for Transliterated Data

Consideration should be given to reviving the approach used for describing non-Latin script items on old catalog cards before the implementation of MARC. In the pre-MARC era, a printed card for a non-Latin script item contained all the information in the original script, with access points and title proper also presented in transliteration. In the context of the BIBFRAME environment, descriptive metadata would be provided in the original script in Instances, and access points and title proper would be provided in transliteration in Works. In fact, BIBFRAME testers have already been following this workflow during testing.

Recording descriptive metadata in non-Latin script in the Instances simplifies the creation and structure of bibliographic description, saving resources in keying transliterated data especially for the many languages which do not have transliteration tools. More importantly, this approach helps meet the needs of non-Latin script users in the increasingly global library ecosystem. In addition, providing access points and title proper in transliteration ensures the organization and arrangement of library materials, and the maintenance of authorized access points.

Generating Transliterated Data to Facilitate BIBFRAME to MARC Conversion

An important workflow step is the conversion of BIBFRAME data to MARC. For the BIBFRAME to MARC conversion process, the generation of paired transliterated fields for descriptive metadata is critical to fulfill two requirements, 1) to align with the current PCC policy for creating bibliographic records in multiple scripts, and 2) to meet the needs of Cataloging Distribution Services (CDS) customers who expect to receive records from Library of Congress containing both Latin and non-Latin scripts for ingestion into their local systems. Library's CDS distributes cataloging data to customers throughout the country. Note: regarding 2), the impact on CDS customers should be verified.

Incorporating ScriptShifter into BIBFRAME

ScriptShifter is being developed to provide automated transliteration of original scripts. This has been a collaborative effort involving a contractor and ABA staff, with contributions from external institutions. Right now, ScriptShifter is capable of handling transliteration in Arabic, Armenian, Chinese, Georgian, Hebrew, Korean. Mongolian, and most of Cyrillic languages. The end goal of ScriptShifter's development is to provide automated transliteration for as many languages as possible. The success of the development of ScriptShifter will significantly streamline the input of non-Latin script materials in BIBFRAME.

Solutions Needed beyond ScriptShifter in Generating Transliterated Data

Currently, materials in many other non-Latin script languages acquired by the Library for its collection do not have an automated tool for transliteration. These languages include South and Southeast Asian languages, Japanese, Tibetan, etc. The question is how best to handle the situation where automated transliteration tools are not available for the languages mentioned earlier. One option is to consider the approach used by OCLC, where angle brackets are used as a placeholder in the empty fields where transliteration would reside. Are there other alternative approaches we can explore?

Conclusion

To summarize, decisions or best practices should be made in the following two areas. First, a clear policy is needed for how much transliterated data we should include in bibliographic description in the BIBFRAME Editor. Second, an approach or strategy should be identified to address the issue of transliteration in cases where automated transliteration tool is not available to facilitate BIBFRAME to MARC conversion.

A timeline for a working plan should be established. Once policy recommendations are identified and formulated, they should be shared with all stakeholders including ABA staff, CDMS, Collections Discovery Group (CDG), CDS, and PCC for comments and feedback. Testing should be done before the official adoption of the recommendations.

Note: Due to the different workflows, authority records and CONSER records are out of scope of this report.