

Report on Non-Latin Script Cross-reference Coding Practice in NACO Name Authority File

By PCC Standing Committee on Applications

September 7, 2022

Introduction

In March 2022, Library of Congress (LC) requested PCC Standing Committee on Applications (SCA) to investigate whether the the abolishment of the practice of coding 008/29 as “b” and adding a 667 note “Non-Latin reference not evaluated” to name authority records (NARs) that contain non-Latin script reference(s) would inadvertently cause library systems changing non-Latin access points in bibliographic records into their authorized forms in Latin script.

History

LC announced the plan to add non-Latin script references into NARs in October 2007. Instead of following [Model A](#) for multiscript records as done in bibliographic records, LC decided to adopt [Model B](#) for the NACO Name Authority File (NAF)¹. As a result, non-Latin script references are recorded in MARC 4XX in NARs. In order to prevent “flipping” of non-Latin script access points recorded in bibliographic records by any integrated library systems (ILS), LC instructed NACO catalogers to:

- code MARC 008/29 (Reference evaluation) as “b” (Tracings are not necessarily consistent with the heading), and
- record a MARC 667 note saying “Non-Latin script reference not evaluated” or “Non-Latin script references not evaluated”

when a NAR contains non-Latin script reference(s). Through data-mining non-Latin bibliographic access point fields in WorldCat, OCLC pre-populated the personal and corporate name authority records in NAF with non-Latin MARC 4XX references in 2008. Besides coding MARC 008/29 as “b” and adding the MARC 667 note, these machine-processed NARs have an additional MARC 667 note “Machine-derived non-Latin script reference project” added.

¹ Non-Latin Script Data in Name Authority Records: Frequently Asked Questions. Available at <https://www.loc.gov/catdir/cpsa/nonlatinfaq.html>

Recent Development

In the past two years, the Chinese-Japanese-Korean (CJK) cataloging community has been working on a project to clean up (aka correct and/or remove invalid) non-Latin script references in those machine-processed personal name authority records. Since project catalogers manually check and remove invalid non-Latin script references, they've asked LC whether they should keep the MARC 008/29 unchanged and leave the MARC 667 note intact. Subsequently, SCA was asked to investigate any potential implications of abolishing this long-standing practice.

Methodology

Upon receiving LC's request, SCA compiled a list of ILS and authority control vendors, primarily based on Marshall Breeding's list of Library Technology Vendors, and emailed those vendors asking if the change would adversely impact data integrity of MARC bibliographic records. SCA reached out to 29 vendors (representing 36 different ILS products and authority processing services) and received 25 responses.

Results

None of the respondents believe the abolishment of the practice would cause unintended "flipping" of non-Latin script access points in bibliographic records. All respondents indicated that their systems or processes either 1) do not touch non-Latin script access points in bibliographic records, or 2) do not take the MARC 008/29 and/or MARC 667 note into account during authority processing. One authority control vendor further clarified that they do not make any distinction between Latin and non-Latin script references in NARs.

Discussions

Based on responses from ILS and authority control vendors, it is safe to say that the abolishment of the special coding practice in NARs with non-Latin script references will not "flip" the non-Latin script access points stored in MARC 880s of bibliographic records. However, there may be edge cases where the change in practice might potentially cause issues.

An authority control vendor stated that they do not process non-Latin script access points in MARC 880. Based on the assumption that no non-Latin script access points are recorded in any regular MARC fields (1XX/6XX/7XX/800-830), it is unnecessary for vendors to differentiate between Latin and non-Latin script references in NARs and handle them separately. As a result, if a non-Latin script access point is recorded in a regular MARC field and matches a non-Latin script reference in the NAF, their authority control process will change that access point into the corresponding Latin script authorized access point causing the loss of non-Latin script data. Any deviation from the existing PCC's Model A multiscript record practice², which records non-Latin script and Latin script access points in MARC 880 and regular MARC fields respectively, will potentially cause a problem.

A related scenario is for libraries operating in a multilingual or non-Latin script environment where they consider the non-Latin script data to be primary. Such libraries may decide to record non-Latin script data in regular MARC fields and the Latin script data in MARC 880s. This would be the exact opposite of current PCC's Model A practice. The Hong Kong Academic Library Link (HKALL), a consortial catalog of higher education institutions in Hong Kong, is an example of such practice. They record non-Latin script data in regular MARC fields and transliteration (aka Latin script data) in MARC 880s for their CJK materials. Besides contributing to the NAF as NACO members, member libraries of HKALL maintain their own Chinese name authority file called [HKCAN](#) simultaneously. According to a cataloger from one of the member libraries, their authority control process currently matches access points against the HKCAN file first, then matches against the NAF when the previous match fails. If a non-Latin access point in a bibliographic record matches a 4XX in the NAF, a link will be maintained in their Alma system between that access point and the matched NAR; however, their Alma system is set up not to flip the non-Latin script access point into its corresponding Latin script authorized access point in this case. Since there is a mechanism already in place to prevent the flip, the HKALL cataloger doesn't think the abolishment of MARC 008/29 coding practice and the 667 note is going to have any detrimental effects on their consortial catalog.

Conclusion

Based on the above survey results and discussions, SCA recommends abolishing the special coding practice for MARC 008/29 and 667 note in NARs that contain non-Latin script cross-references.

² MARC 21 Bibliographic Format [Appendix D](#) does not dictate whether non-Latin script data to be recorded in regular MARC fields or 880 fields. It is the customary practice for English-speaking libraries to record non-Latin script data in 880s