Program for Cooperative Cataloging (PCC) Standing Committee on Standards (SCS) Task Group on Developing Guidelines for Using ISO 639-3 Language Codes for Library Cataloging

Interim Report August 15 2022

Introduction

The Task Group on Developing Guidelines for Using ISO 639-3 Language Codes for Library Cataloging members include Kelley McGrath, Christopher Dillon, Greta Heng, TJ Kao, Charlene Morrison, Robert Rendall, Charles Riley, and Lana Soglasnova. This group was charged to develop a set of guidelines for when and how to use ISO 639-3 language codes in library cataloging and expanding the guidelines developed for the BABEL test.

Per the charge laid out, this interim report will describe the work done so far. It will also include the tasks the group are taking on now and issues surfaced along the way.

General Report

The group began with reviewing work done by the previous two groups and analyzing the test survey results gathered between May and December 2021. The group examined each response and assigned a category to it, and divided the categories into those that require further actions and those that do not. Among the former categories, the group identified several major topics that require more consideration and policy decisions, which are listed below.

The task group expects that a revision of the guidelines issued for the test will be needed to support PCC use of ISO 693-3 in the future, with modifications reflecting decisions made related to the issues discussed in this report. For our current cataloging environment, the group expects to recommend that MARC language codes should continue to be required for PCC records, and that ISO 639-3 codes should optionally be added by catalogers whenever they feel those codes will add information to the description, in other words when they identify a language more specifically than the MARC codes do. The only other major change that the task group anticipates making to the test guidelines is removing the recommendation to use the ISO 639-3 code "und" for undetermined when the language cannot be identified with the granularity required by ISO 639-3 but a general language group is known. Instead, the task group will recommend using the MARC language group code only in this situation. The task group is also developing an FAQ as a parallel method of providing guidance.

Topics under Discussion

Languages vs. dialects (scope of ISO 639-3)

ISO 639-3 contains separate codes for everything that its Registration Authority (the Summer Institute of Linguistics) considers a separate language. It does not contain codes for dialects within those languages. Their language vs. dialect decisions may differ from the cataloger's sense of the situation, or from how the language is presented in the resource being cataloged. Research may be necessary to determine the

correct code to use, and the lack of cross-references in the official source for the codes will make this particularly difficult in some cases.

When the resource cataloged is in a linguistic form that SIL considers to be a dialect, the code for the larger language should be used. Information about the specific variety or "dialect" can be entered in a 546 note.

Other vocabularies are available that do cover regional variants (IETF BCP-47) and scripts (ISO 15924). The task group has determined that these vocabularies are too complex for incorporation into MARC, but it may become possible to use them with BIBFRAME in the future.

Historical forms of languages

ISO 639-3 considers many earlier forms of modern languages to be separate languages, with separate codes. So for example Old English (ca. 450-1100) (*ang*) and Middle English (1100-1500) (*enm*) have their own codes, which should be used whenever appropriate. If no separate codes are available for earlier forms, as for example in the case of Portuguese, the code for the modern language (*por*) should be used for all periods. Whenever needed, additional information about the language can be recorded in a 546 note.

The ISO 639-3 code name sometimes indicates the approximate time period covered by the historical language code, and sometimes does not. If the language label does not contain a date range, it may be necessary for the cataloger to do some research or consult community guidelines to determine whether a resource produced in a particular time period should be described with a code for a historical language. When appropriate, PCC should encourage expert language communities to develop such guidelines.

Granularity

When assigning language codes from ISO 639-3, the task group recommends assigning the most specific language code that is applicable to the resource. In some cases, catalogers will not be able to identify the language of the resource with the granularity required by ISO 639-3. While MARC language codes include codes for certain groups of languages, ISO 639-3 does not include any codes for groups of languages. When a cataloger has entered a MARC group code but is unable to identify a specific language code within ISO 639-3, the cataloger should omit the ISO 639-3 code and use the MARC group language code only. Thus the workflow would be:

> Step 1. Assign a MARC language code as usual (always required; may be an individual language code or a group code).
> Step 2. Optionally assign a 639-3 code (which will always be an individual language code) if the correct code can be identified; otherwise omit.

This recommendation should be reviewed in the future and updated if we move to an environment where the MARC language codes are not used.

Macrolanguages

Our recommendation for guidance on using macrolanguage codes continues to be that the general macrolanguage code be applied only in cases where a more specific language cannot be reliably identified by the cataloger if a macrolanguage code is available. Thus, for example, [gba] would be suitable for identifying material as being in Gbaya only if it is not known more specifically whether it is in Northwest Gbaya [gya], Southwest Gbaya [gso], Gbaya-Mbodomo [gmm], Gbaya-Bozoum [gbq], Gbaya-Bossangoa [gbp], or Bokoto [bdt].

Looking up and identifying codes

Participants in the BABEL test experienced difficulties looking up and identifying the correct ISO 639-3 code. The previous task group recommended using the official SIL site to look up the language codes. However, cross-references are not provided and, since results include codes from the entire ISO 639 family of language codes, participants sometimes became confused about which vocabulary they were looking at. LC and SIL are working together to post the ISO 639-3 vocabulary on id.loc.gov, but it is not known how long this process will take. If LC is unable to post this vocabulary in the near future, the task group recommends investigating the possibility of using the PCC's proposed Wikibase instance to provide better access.

Mapping issues

The mapping of ISO 639-2 to 639-3 depends in part on language classification decisions that are not themselves uncontroversial. Where possible, a mapping based on consensus over classification has been sought. Some of the classifications that relate to language families and their coding in ISO 639-2 are problematic, such as Finno-Ugrian [fiu] and Altaic [tut]. The mapping tends toward affiliating languages with the narrowest subgroup to which they belong, although this has not been strictly followed. Mappings of Roma languages, Malay trade and creole languages, and historical languages each pose unique challenges. As new research and debate comes to light, it is understood that mappings may need to be revisited, and that the mapping is at best a heuristic device.

Topics Outside the Task Group's Charge

Involvement of specialist cataloging communities

The task group is not able to predict all the specific questions that may come up in the application of these codes for individual languages, and recommends that PCC should include language expert communities in the process of adopting ISO 639-3. These communities will have the knowledge and

expertise to answer language-specific questions and establish any more specific best practices or guidelines that may be needed. The Task Group is compiling a contact list of groups with such language expertise. So far, the following groups have been identified as potential partners, but the task group welcomes additional recommendations.

- ALA ACRL European Studies Section Cataloging Discussion Group
- ALA ACRL Language and Linguistics Discussion Group
- ALA Core Committee on Cataloging Asian & African Materials Committee
- American Indian Library Association
- Australian Institute of Aboriginal and Torres Strait Islander Studies
- CJK NACO Project
- Committee on Research Materials on Southeast Asia (CORMOSEA) - Technical processing
- Committee on South Asian Libraries and Documentation (Consald)
- Council on East Asian Libraries (CEAL) Committee on Technical Processing (CTP)
- Middle East Librarians Association (MELA) Committee on Cataloging
- National Committee for Information Resources on Asia (U.K.)
- Slavic Cataloging Manual

Potential future use of ISO 639-3 only

The guidelines developed by the Task Group are intended for use in our current MARC environment, and assume continued use of the MARC language codes, supplemented by the ISO 639-3 codes in situations where they allow catalogers to record more specific information. In future non-MARC environments, it is possible that PCC will want to use ISO 639-3 codes only, for consistency. However, this will present additional challenges because of the difficulty of determining the correct specific 639-3 code in some cases as discussed above under Granularity. Additional policy decisions and guidance for catalogers would be required to support the use of ISO 639-3 alone.