



Non-Latin Scripts Cataloging in Sinopia

Larisa Walsh and Patricia Thurston (LD4 Non-Latin Script Materials Affinity Group)

PCC Sinopia Cataloging Affinity group meeting

(December 16, 2021)

The LD4P2 Non-Latin Script Materials Affinity Group

- Creation of many metadata description sets for non-Latin resources
- Identification of some issues
- Discussions about how non-Latin script data could be used and represented in Sinopia
- “Common Practices and Recommendations” – summary of best practices in native scripts cataloging in Sinopia, developed by the members of the group, along with issues and possible solutions how to resolve them

What non-Latin cataloging community wants from linked data editor?

Use cases

Top priority

- Enter native script data in Sinopia (all scripts)
- Copy native script data to and from external sources into Sinopia
- Enter and display diacritics
- Provide transliteration for native script data
- Link native script data fields with corresponding romanized data fields

Desired features

- Exhibit correct directionality and alignment for right to left scripts
- Exhibit correct directionality for vertical scripts
- Create authority records for non-Latin names
- Provide automatic transliteration from native script to Latin and Latin to native script

Enter native script data in Sinopia (all scripts)

- Sinopia handles well the input, search and display of the more-highly used native script characters. Native scripts can be entered in Sinopia by:
- 1 - copy-and-paste into Sinopia from external resource (e.g. OCLC Connexion, Google Translate, Word documents, various Internet tools, etc.)
- 2 – entering native script directly into Sinopia, by using: the Diacritic Picker, the Google Input Tool from the Chrome browser or, Windows on-screen keyboard
- Challenge: when copying from external sources is not feasible, using input tools or Diacritic Picker is not the best solution – too time consuming

Transliteration: Diacritic Picker

▼ Main Title [↗](#)

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>

Enter a literal

 ä 

- Marked as a letter “a” with umlaut in the right corner of each descriptive element in resource templates
- Provides characters for vernacular scripts as well as Latin script characters that include diacritics
- Not a set of diacritics, but a set of letters combined with a diacritic - precomposed
- Currently includes characters for 19 scripts:

• Latin	Arabic	Sinhala
• IPA	Arabic Extended	Devanagari
• Symbols	Hebrew	Gujarati
• Greek	Bengali	Thai
• Greek extended	Tamil	Lao
• Cyrillic	Telugu	Khmer
		Canadian Aboriginal

Diacritic Picker

▼ Main Title [↗](#)

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>

Enter a literal

Вой ä en ✖

Latin
IPA
Symbols
Greek
Greek Extended
Cyrillic
Arabic
Arabic Extended
Hebrew
Bengali

È è É é Ê ë Ë ë Ì ì Í í Î î Ï ï Ñ ñ Ò ò Ó ó Ô ô Õ õ Ö ö ×

Ж ж З з Ѕ ѕ Ѓ ѓ Ё ё Ї ї Њ њ Ќ ќ Ћ ћ Ь ъ Ї й Й й Ы ы Ь ъ Я я

Ј ј К к Ќ ќ Ѓ ѓ Є ф Ѕ ѕ І ф Ї й Ы ы Ь ъ Ь ъ Q q Л л

< >

Diacritic picker (2) -

- Several very common scripts are missing: Chinese, Japanese and Korean.
- Also missing: Adlam, Armenian, Balinese, Burmese, Cham, Cherokee, Coptic, Deseret, Ethiopic, Georgian, Gurmukhi, Javanese, Kannada, Kayah Li, Lepcha, Malayalam, Meetei Mayek, Mongolian, N'ko, Oriya, Pahawh Hmong, Panjabi, Phags-Pa, Shavian, Syriac, Tibetan, Tifinagh, Vai
- Some characters are still missing from existing tables
- Users must sometimes click through multiple tables to get the desired character. For example: To find diacritics used in Arabic the user must search through the sets for Arabic, Arabic Extended, Latin, Latin Extended, and Symbols, to find the needed characters.
- Many compound characters are included in the IPA set (International Phonetic Alphabet), the set is incomplete

Diacritic Picker (3)

Desired functions:

- Arrange the filter list by language, not by script. This would eliminate the navigation problems that come with going between several script tables to find a character
- Within right-to-left scripts lists, arrange the characters in alphabetical order, with the first letter beginning on the right, and subsequent letters going out to the left.
- Create a mouse-over prompt with the name of the character or symbol.

Recommended practice:

Use Diacritic Picker for inserting single letters with diacritics or a relatively short text. For lengthier text passages use the copy-and-paste to move native script data created in external sources into Sinopia

Recording Initial Articles

- Handled differently in Sinopia by cohort institutions
- **Recommended practice:** Leave out the initial articles in the *Work* description for both romanized and native script titles. Some institutions include the initial article in the Variant Work Title in a Work description.

Work Title

+ Add another

Class: <http://id.loc.gov/ontologies/bibframe/Title>

▼ Preferred Title for Work *

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>

Enter a literal

al-Taghārīd al-Bulbulīyah bi-ba'd ittisālātī bi-al-Awā'il al-Sunbulīyah

ä ar-Latn-t-ar-m0-alaloc

Work Title Variation

+ Add another

Class: <http://id.loc.gov/ontologies/bibframe/VariantTitle>

▼ Variant Title for Work

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>

Enter a literal

al-Taghārīd al-Bulbulīyah bi-ba'd ittisālātī bi-al-Awā'il al-Sunbulīyah

ä

Include initial articles in the *Instance* title. Add *Variant Instance* title without an article.

Transliteration in Sinopia:

What data elements should be Romanized and why

- Searching by title (romanized or native script) in Sinopia brings together both the Work title (romanized) and Instance title (in native script) **only** when Work and Instance description sets include both (transliterated and native script titles), or linked by URIs through Instance of, or Has BIBFRAME Instance elements.
- Adding a variant title in native script to the Work template makes the retrieval more effective. The same is true for other elements of description that currently only have the option to be entered in native script: Statement of responsibility, Edition statement, Transcribed provider statement, Series statement, Contents note.

Statement on Transliteration

- The group strongly supports continuing the practice of adding romanized data to descriptive data elements. These data elements are currently missing in Sinopia in several places in the Instance Template: Transcribed Provider Statement, Statement of Responsibility, Series statement, Edition statement and Contents note.
- [Survey on Romanization](#) results indicated that the library and research community still heavily depend on romanization for the majority of library operations and research. The addition of romanized data in the descriptive elements continues to be valued as a means of providing greater discoverability of the library holdings, thus supporting access to library collections.

Transcription of Dates in Non-Gregorian Calendar

- Desire to transcribe the dates exactly as they appear on the piece while making them machine actionable.
- **Recommended practice:** For a non-Gregorian date that appears on the piece enter that date in the *Transcribed Provider Statement (in Latin script)* and in the separate *Date* element . Optionally, add the corresponding date of the Gregorian calendar in square brackets, after the non-Gregorian date. With the new property “Add another Transcribed Provider Statement” it is possible to add a date in Gregorian calendar, or in native script. Add a note indicating the form or calendar in which the date appears on the

▼ Date

Property: <http://id.loc.gov/ontologies/bibframe/date>

Enter a literal

Minguo 61

ä zh-Latn

Property: <http://id.loc.gov/ontologies/bibframe/date>

Enter a literal

1972

ä en-Latn

Property: <http://id.loc.gov/ontologies/bibframe/date>

Enter a literal

民國61

ä zh-Hans-t-zh-m0-pinyin

+ Add another Date

Desired improvements – transliteration

- “Add a **subtag for the script** for descriptive elements. When cataloging materials in native scripts, it is quite common to have a title or a publisher’s statement in a different script than the text of the content”. (DONE!)
- “Add a **subtag for transliteration schema**: for example, when Chinese characters romanized per Pinyin or Wade-Giles standards. For written Chinese, there are two different types of scripts, simplified Chinese and traditional Chinese. It is essential to identify the scripts a resource is printed in. At this moment, no language code is sufficient for this purpose”. (DONE!)
- Adding a **version number**, or a date to the **romanization schema** would be desirable. ALA/LC Romanization Tables, for example, have an older, pre-1997 version. Example: Khmer word នឹង is found in LC's catalog as both "niñ" (pre-1997 version) and "ning" (latest version)] –(Not implemented yet)

Desired improvements – transliteration (2)

- ★ Develop the functionality to create automatic transliteration, similar to the [OCLC transliteration macros created by Joel Hahn](#)
- **Mixed languages and scripts:** issue remains open, but if there could be a way to indicate what language/script is part of the data (for example, some words in the title appear in one language or script, and others appear in a different script (or language)).

Desired improvements - Allow for Bi-directional Script Data in a Single Field (as can be done in MARC)

- It is not possible to enter bidirectional text in a single field in Sinopia. Text in a right-to-left script must be entered in its own field (with a language value assigned).

▼ Main Title [🔗](#)

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>

Enter a literal

نوميديا وروما الامبراطورية

ä ar-Arab-t-ar-m0-buckwalt

245	1	0	نوميديا وروما الامبراطورية : b تحولات اقتصادية واجتماعية في ظل الاحتلال / c ا.د. محمد البشير شنيطي, استاذ الآثار والحضارات القديمة بجامعة الجزائر
245	1	0	Nūmidīyā wa-Rūmā al-imbrātūrīyah : b taḥawwulāt iqtisādīyah wa-ijtimā'iyah fī ḡill al-iḥtilāl / c U.D. Muḥammad al-Bashīr Shinitī, ustādh al-āthār wa-al-ḥadārāt al-qadīmah bi-Jāmi'at al-Jazā'ir.

Desired improvements (3)

- Exhibit correct directionality and alignment for right to left scripts
- Exhibit correct directionality for vertical scripts
- Create authority records for non-Latin names
- Provide automatic transliteration from native script to Latin and Latin to native script

**Thank you
very much for
your attention!**

