Cover Sheet

Draft-PCC Guidelines for Creating Bibliographic Records in Multiple Character sets.

and

Preliminary Report: Task Force on Non-Latin Script Cataloging Documentation.

The accompanying draft guidelines are preliminary and not complete. In particular, the section for special guidelines for particular languages is not complete. Further comments are requested from PCC in order for the task group to receive input and advice for completing those sections and possibly revising the general section (see Specific comments requested below). The draft guidelines, if adopted, would establish a standard practice for PCC catalogers adding parallel non-Latin data to MARC bibliographic records. The final draft is due Dec. 2009.

The accompanying Preliminary report includes the task force roster, background information, scope and organization of guidelines draft, rationale for the draft guidelines, and issues that were brought up in the task group during discussion.

Specific comments requested:

- On the efficacy and organization of the draft: having a general section applying to all scripts, and having a separate section with special rules for particular scripts and languages as necessary
- On guidelines for headings (especially for right-to-left scripts (e.g., Hebrew, Arabic, etc.)):
    - allowing for the omission of the use of qualifiers
    - or, allowing the use of non-Latin data in qualifiers
    - or, allowing non-Latin parallel data for headings not established according to ALA/LC transliteration tables
- On the pros and cons of recording non-Latin data for headings only in references in authority records (i.e., no non-Latin parallel data for headings in bibliographic records)

Submitted Sept. 1, 2009

Deadline for comments: Oct. 7, 2009

Preliminary Report: Task Force on Non-Latin Script Cataloging Documentation
08/17/2009

To: PCC, Standing Committee on Standards

Re: Task Force on Non-Latin Script Cataloging Documentation Preliminary report

This preliminary report consists of the report itself containing background, recommendations, and issues for discussion, plus a Draft-PCC Guidelines for Creating Bibliographic Records in Multiple Character sets.

Task force members:

Benjamin Abrahamse, MIT (Arabic, Hebrew)
Nora Avetyan, UCLA (Persian, some Greek)
Sharon Benamou UCLA (Hebrew, Yiddish), began Sept. 2009
Diana Brooking, Univ. of Washington (Cyrillic)
Shi Deng, UC San Diego (CJK)
Joanna Dyla, Stanford (Cyrillic)
Peter Fletcher, UCLA, chair, PCC, SCS rep. (Cyrillic)
Caroline Miller, UCLA (Hebrew/Yiddish)
Robert Rendall, Columbia (Cyrillic, some ability "nonexpert" with Greek, Arabic, Chinese)
Dave Reser, LC liaison
Julie Su, San Diego State (CJK)
Keiko Suzuki, Yale (CJK)
Patricia Thurston, Yale (Cyrillic)
Jiping Wu, Coop Section (formerly Coop Team) Library of Congress (CJK)

Background

Following up on the recommendation 3 of the *Task Force on Non-English Access*. Rev. Mar 16, 2007, (http://www.ala.org/ala/mgrps/divs/alcts/ianda/nonenglish/07marchrpt.pdf) "to review and update the core level supplement on "Guidelines for Multiple Character Sets" (http://www.loc.gov/catdir/pcc/bibco/coreintro.html#9)," the PCC Standing Committee on Standards revised the document to bring it up to date with current developments. However, the committee considered that the creation of more comprehensive documentation would be helpful to PCC catalogers dealing with non-Latin materials and would be more in the spirit of the Task Force's recommendation. A review of current PCC documentation (and some non-PCC documents) on non-Latin script cataloging conducted during the revision of the above document revealed some inconsistencies in practice among current documents on the one hand, and a lack or obsolescence of documentation for some scripts altogether. This leaves catalogers to extrapolate from documentation dealing with scripts other than the ones they are working with and, where there is no documentation for a given script, allows for inconsistencies among PCC bibliographic record practices. Upon recommendations made during the May 2008 BIBCO/CONSER Operations Committee meetings, the PCC has decided to form a task force to produce needed non-Latin cataloging documentation for PCC bibliographic records.

Documents reviewed:

BIBCO core record standards. 9. Guidelines for multiple character sets. [Updated 2008]

CONSER Editing Guide. 1994 ed. Section F, appendixes: appendix E. Creating records with data in non-Roman script for Arabic serials [2003]. Appendix O. Creating records with data in nonroman script for Chinese, Japanese, and Korean serials [updated Sept. 2008].

Draft: creating records with data in non-Latin scripts for Cyrillic (Russian) monographs. Rev. July 22, 2008 [internal Library of Congress Document]

White paper: issues related to non-Latin characters in name authority records. Rev. draft, Dec. 19, 2007.

Non-Roman Core Record Task Group. Final report. [PCC Web site; no date given]

Organization

The draft guidelines document has a general section that is followed by as section for specific script guidelines that is also"under development". We hope that the organization of the document proves to be easy to use.

Rationale of draft document

We endeavored to create a general section of guidelines applicable to all scripts/languages, with the hope of creating a document that is convenient to use and to minimize the necessity and scope of future revisions as new languages and scripts become available for use in the changing operating environment. Guidelines for new scripts/languages can be added as necessary if the general guidelines would not otherwise cover the unique requirements of a particular language.

We have also attempted to remain operating environment neutral, but OCLC is mentioned since PCC libraries use it for most cataloging. Also we do not want to duplicate OCLC or other systems documentation

Optional v mandatory fields: descriptive fields are designated "mandatory if applicable" because they are transcribed from the resource. Notes are optional since most notes in AACR/MARC cataloging are optional. Headings fields are optional since some libraries may prefer to utilize or provide non-Latin data only as references in authority records.

Scope and policy: This draft applies only to bibliographic record practice. Also, the policy and practice in place for the *CONSER editing guide* appendices E. and O. have been adhered to. However, when these rules do not conform to widespread current cataloging practice for a particular script, special rules or exceptions to the CEG appendices could be developed in the section(s) for special rules (see Issues 1. below and section 2.5 of the draft guidelines).

Discussion Issues

1. Possible exceptions to the general standards (see section 2.5 of draft guidelines): provide parallel non-Latin script data for headings not established according to ALA/LC standard transliteration; provide also non-Latin data in cataloger supplied qualifiers (whether or not the qualifier has received standard romanization) or omit the qualifier in the parallel heading field. Current practice for Hebraica and some other right-to-left script monograph PCC and LC catalogers varies widely in the provision of non-Latin parallel fields. These guidelines could allow for such exceptions in the section(s) for specific scripts/languages, the rationale being that the guidelines could at least achieve standardization within a given community, and it is doubtful that if we recommended strict adherence to the CEG appendices that those monographic cataloging communities would stop adding this non-standard data.

2. Some members of the task force would prefer that non-Latin data for headings reside only as references in authority records rather than as parallel fields in bibliographic records. This draft document maintains the status quo in that it is optional to provide non-Latin parallel data for headings fields in bibliographic

records. However, it should be also noted that this approach—allowing non-Latin variants to be recorded only in the authority records—could be a possible solution to Issue 1 above.

Recording non-Latin variants only in the authority record:

- Pro: non-Latin authority record references refer researchers to the established forms of headings; less work in bibliographic records (heading is created once in AR); less opportunity for proliferation of multiple variant forms and practices of parallel data in headings in bibliographic records.
- Con: some systems do not utilize authority records effectively or at all for search and retrieval; non-Latin parallel heading fields also serve a display purpose in addition to access (it should be pointed out that some developments such as the CONSER standard record, the statement of responsibility is not required, thus the heading field is the only place a user unfamiliar with romanization may see an author's name in non-Latin script; non-Latin forms in bibliographic records can be used to pre-populate references in authority records; bib records often exist outside of the context of the catalog supported by authorities, e.g., new book lists, outputs to bibliography software, webharvesting, etc.; for CJK records where a much larger percentage of authority records are "non-unique" because different CJK characters romanize to the same form—these records can have many unique non-Latin references that actually represent different people—if you don't allow the non-Latin form in the bib record, you can't assist users in identifying the particular person involved.

Draft-PCC Guidelines for Creating Bibliographic Records in Multiple Character sets

Aug. 27, 2009

This document contains guidelines for providing non-Latin script fields that are parallel to their Latin (Romanized[1]) counterparts within MARC bibliographic records using MARC21 Model A (original script and transliteration)[2]. These guidelines are applicable to operating environments where current AACR/RDA-MARC cataloging is input. However, references are made to OCLC, given its importance as an operating environment and as a union cataloging for most PCC libraries.

The guidelines apply to the scripts/languages which are represented by the current MARC-8 repertoire of UTF-8 (Chinese, Japanese, Korean, Hebrew, Yiddish, Arabic, Persian, Greek, Cyrillic). However, PCC-BIBCO (not NACO, SACO, or CONSER) records can currently accommodate scripts that can be entered in OCLC but are outside of the MARC-8 Repertoire. (Please refer to appropriate OCLC systems documentation). As more scripts/languages are added in the future, this document will be updated as necessary.

Terms used:

Non-Latin script, non-Roman script or *original script:* characters written in their original alphabet

Romanization: conversion of names or text not written in the roman (Latin) alphabet to Roman/Latin-alphabet form (based on AACR2 Glossary)

Transliteration: conversion of text from one type of alphabet to another

Introduction

Although the decision to include data in non-Latin form in any PCC record is strictly optional, when that option is exercised, it must be done so according to these guidelines. Catalogers adding data in non-Latin form to records are encouraged to consider the future use and international implications of their records and to include as much of the original script data as necessary to facilitate the identification of and location of this often scarce material and its component parts. Many library and bibliographic search systems now have non-Latin searching capabilities.

In all cases PCC libraries must include in Latin form all of the data elements as specified in the standard for the type of material being cataloged.

Aside from preliminaries (Introduction, Record characteristics), the document is divided into two main parts: 2. *General guidelines* that broadly apply to all language/script groups; and, 3. *Special language/script-specific guidelines* for problems unique to a language or script.

1. Record characteristics

If original script is used in a bibliographic record, all descriptive fields listed below as MA (mandatory if applicable) must appear in parallel romanized and original-script fields.

All headings must be established in Latin script according to AACR2, regarless of the presence or absence of non-Latin script elsewhere in the record.

MARC 880 fields contain the parallel original script data. However, in the OCLC operating environment, parallel fields display as the same MARC tags and indicators as their Latin equivalents and are linked, or paired, to them. (with the exception of the 600-630, 651 indicators as described in 2.5 below). In addition, field 066 is system-supplied and identifies the non-Latin script character sets present in the record. Please refer to relevant system or MARC documentation for more information.

Macros and "automatic" transliteration: Given the current limitations of scripting technology, macros whose purpose is to automate the transliteration process should be deployed with care by cataloging agencies, and the results should be subject to review by catalogers who possess expert knowledge of the script and language in question

2. General guidelines:

With some restrictions indicated here, parallel non-Latin data may be created for those fields, parts of fields, or subfields that are Romanized according to the ALA-LC Romanization tables[1].

Mandatory (if applicable) (MA)	Optional
245 (Title/statement of responsibility)	1XX (Main entry)
246 (Varying form of title) 247 (Former title)	240 (Uniform title)
250 (Edition statement)	362 (Dates of publication and/or sequential designation)
260 (Publication, distribution, etc.)	5XX (Notes)

490 (Series statements)	600-630, 651 (Subject access fields)
740 (Title added entry)	700-730 (Added entries)
	760-787 (Linking entries)
	8XX (Series added entries)

2.1 *Fixed Field:* Code the fixed field Modified Record Code (008/38) # (blank).

2.2 Fields supplied by ISSN centers: 210 (abbreviated title), 222 (key title): do not attempt to provide non-Latin data for these fields.

2.3. Descriptive fields (MA) (245, 246, 247, 250, 260, 490, 740)

For these descriptive fields, parallel non-Latin script data is given only for those elements that *appear in non-Latin script transcribed from the item itself*.

Examples:

245 00 ‡a Родное и вселенское : ‡b к 60-летию Николая Николаевича Лисового / ‡c [ответственный редактор, А.В. Назаренко].
245 00 ‡a Rodnoe i vselenskoe : ‡b k 60-letii͡u͡ Nikolai͡a͡ Nikolaevicha Lisovogo / ‡c [otvetstvennyĭ redaktor, A.V. Nazarenko].

245 00 ‡a פייערלעכער אוועגט און בינע פארשטעלונג צום דערשיינען פון דוד עדעלשטאט געדענק־בוך.

245 00 ‡a Fayerlekher oṿenṭ un bine forshṭelung tsum dershaynen fun Doṿid Edelshṭaṭ Gedenḳ-bukh.

245 10 ‡a أكتب لكم من الحرملك : ‡b وقالت شهرزاد / ‡c هالة سرحان.

245 10 ‡a Aktubu lakum min al-ḥaramlik : ‡b wa-qālat Shahrazād / ‡c Hālah Sarḥān.

490 1# ‡a 崇文学术文库

490 1# ‡a Chong wen xue shu wen ku

490 0# ‡a سلسلة النصوص القانونية ؛ ‡v العدد 14

490 0# Silsilat al-nuṣūṣ al-qānūnīyah ; ‡v al-‘adad 14

Mixed Script Descriptive Fields*:* for descriptive fields including both Latin and non-Latin data, supply a parallel field transcribing the Latin data as found and including the non-Latin data in its original script.

Examples:

Publication information:

260 ## ‡a Москва : ‡b MediaPress [Publisher name appears in Latin script on the piece]
260 ## ‡a Moskva : ‡b MediaPress

Parallel titles:

245 00 ‡a 香港統計月刊 = ‡b Hong Kong monthly digest of statistics.
245 00 ‡a Xianggang tong ji yue kan = ‡b Hong Kong monthly digest of statistics.

245 02 ‡a الجريدة القضائية = ‡b Le journal judiciaire.
245 03 ‡a al-Jarīdah al-qaḍāʾīyah = ‡b Le journal judiciaire.

2.4 ***Notes (Optional) (5XX, 362)***

Only give non-Latin parallel notes for notes that cite non-Latin data that appear on the item itself. PCC cataloging is based on AACR2 and thus the language of cataloging for the PCC is English[3], therefore do not translate introductory phrases or other cataloger-supplied data into other languages and scripts.

Examples:

362 1# ‡a Began with дек. 1998 г.

362 1# ǂa Began with dek. 1998 g.

500 ## ǂa At head of title: Императорское православное палестинское общество. Научный совет РАН "Роль религий в истории." Паломнический центр Московского патриархата.
500 ## ǂa At head of title: Imperatorskoe pravoslavnoe palestinskoe obshchestvo. Nauchnyĭ sovet RAN "Rolʹ religiĭ v istorii." Palomnicheskiĭ t͡s͡entr Moskovskogo patriarkhata.

550 ## ǂa Official organ of: 한국 관세 연구소.
550 ## ǂa Official organ of: Han’guk Kwanse Yŏn’guso.

580 ## ǂa Split into: دراسات انسانية , and: دراسات قانونية
580 ## ǂa Split into: Dirāsāt insānīyah, and: Dirāsāt qānūnīyah.

500 ǂa Editor: 青木昌彦
500 ǂa Editor: Aoki Masahiko

But not:

500 ǂa 編者: 青木昌彦
500 ǂa Editor: Aoki Masahiko

2.5 Headings (optional) (1XX, 240, 600-630, 651, 700-730, 8XX)

For subjects and headings fields, non-Latin equivalents are given only for headings and parts of headings that have received standard[1] Romanization. Do not supply non-Latin data for cataloger supplied qualifiers in corporate names, geographic names, uniform titles, or headings and parts of headings established in a conventional, Latin-style. (See section 3 for exceptions to this rule)

Do not supply parallel non-Latin data for topical subject headings (MARC tag 650) or geographic subdivisions ($z) used in them. For 600-630, and 651 non-Latin data fields: code the second indicator “4” (Source not specified).

Examples

100 1# ǂa Широкорад, А. Б. ǂq (Александр Борисович)

100 1# ‡a Shirokorad, A. B. ‡q (Aleksandr Borisovich)

110 2# ‡a Институт социальных и общинных работников (Minsk, Belarus)
110 2# ‡a Institut sot´s˜ial′nykh i obshchinnykh rabotnikov (Minsk, Belarus)

111 2# ‡a 当代语言学理论和汉语研究国际学术报告会 ‡d (2006 : ‡c Beijing, China)
111 2# ‡a Dang dai yu yan xue li lun he Han yu yan jiu guo ji xue shu bao gao hui ‡d (2006 : ‡c Beijing, China)

240 10 ‡a Преступление и наказание. ‡l English
240 10 ‡a Prestuplenie i nakazanie. ‡l English

610 24 ‡a 中國共產黨 ‡x History ‡y 20th century.
610 20 ‡a Zhongguo gong chan dang ‡x History ‡y 20th century.

700 1# ‡a 張廣文, ‡d 1949-
700 1# ‡a Zhang, Guangwen, ‡d 1949-

710 2# ‡a 北京电视艺术中心音像出版社.
710 2# ‡a Beijing dian shi yi shu zhong xin yin xiang chu ban she.

710 2# ‡a 中央日報 (Firm)
710 2# ‡a Zhong yang ri bao (Firm)

Examples of headings that have not received standard[1] romanization:

100 0# ‡a John, $c of Kronstadt, Saint, $d 1829-1909.
do not add parallel field:
100 0# ‡a Иоанн, $c of Kronstadt, Saint, $d 1829-1909.
do not add parallel field:
100 0# ‡a Иоанн, $c Кронштадский, Святой, $d 1829-1909.

700 1# ‡a Cheung, Leslie, $d 1956-2003.
do not add parallel field:

700 1# ‡a 張國榮, $d 1956-2003.

2.6 *Linking entries and note (765-787, 580)*

Give parallel non-Latin script fields only when the linked record is a record containing non-Latin script data (and the non-Latin script data are, therefore, readily available) and the record on which the link is being recorded also has non-Latin script data. Likewise, provide a parallel complex linking entry note field (580) only when all entries cited have corresponding non-Latin script records.

Examples

785 00 ‡a Moscow (Russia). Мэрия. ‡t Вестник Мэра и Правительства Москвы ‡w (DLC) 2007222946 ‡w (OCoLC)51887397
785 00 ‡a Moscow (Russia). Mėrii´a˜. ‡t Vestnik Mėra i Pravitel′stva Moskvy ‡w (DLC) 2007222946 ‡w (OCoLC)51887397

780 00 ‡a Nongŏp Kwahak Yŏn'guwŏn (Korea). $t Hakpo - Nongŏp Kwahak Yŏn'guwŏn
[No parallel field; related record is romanized only].

3. Special language or script guidelines *[UNDER CONSTRUCTION]*

3.1 Arabic/Persian

3.2. CJK

3.2.1. For personal names given in a parallel nonroman script field, do not follow a surname with a comma.

3.2.2. Input spaces between lexical units in nonroman fields that consist solely of Korean hangul, or hangul and ideographs, or any ideographs that are romanized as Korean language words, so that the nonroman data mirrors the spacing in the parallel romanized fields (which are romanized according to

the guidelines in the ALA-LC Romanization Tables). Do not input spaces between characters in nonroman data that are in Japanese or Chinese scripts representing Japanese or Chinese words."

Examples

245 00 $a 외국 과학 기술 통보. $p 전자 공학.
245 00 $a Oeguk kwahak kisul t‘ongbo. $p Chŏnja konghak.

245 10 $a 輸銀 年鑑. (Chinese characters, with spacing between words)
245 10 $a Suŭn yŏn’gam. (Korean Romanization)

3.3 Cyrillic

3.3.1 Characters omitted in Romanization

ALA-LC Romanization Tables calls for omitting a terminal hard sign “ъ” in Russian (pre-revolutionary (1917) script). For Bulgarian, it instructs the cataloger to disregard a soft sign “ь” and, at the end of a word, to disregard the character ”ъ”. Thus, whether or not the cataloger uses a macro, manual editing is required in these specific cases.

Examples:

Russian:

245 10 Вѣсти изъ экспедиціи П.К. Козлова : $b изъ писемъ къ секретарю Общества : съ приложеніемъ карты.

245 10 Vi˜e˘sti iz ėkspedit˜s˘īi P.K. Kozlova : $b iz pisem k sekretari˜u˘ Obshchestva : s prilozhenīem karty.

Bulgarian:

245 10 Руска книжовна речь : $b въ образци.
245 10 Ruska knizhovna rech : $b v obrazt˜s˘i.

3.4. Greek

3.5 Hebrew/Yiddish

Appendices:

PCC documentation upon which this document is based

BIBCO core record standards. 9. Guidelines for multiple character sets. [updated 2008]

CONSER Editing Guide. 1994 ed. Section F, appendixes: appendix E. Creating records with data in non-Roman script for Arabic serials [2003]. appendix O. Creating records with data in nonroman script for Chinese, Japanese, and Korean serials [updated Sept. 2008].

---

[1] Standards for Romanization: ALA-LC Romanization Tables: Transliteration Schemes for Non-Roman Scripts (1997 ed.), Washington: Library of Congress, Cataloging Distribution Service, 1997. Specific scripts may also have further documentation regarding Romanization.

[2] MARC21 format for bibliographic data. App. D. Multiscript records

[3] Except for French language records create by Library and Archives Canada