

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

[Revised for RDA, June 28, 2016; last revised: September 7, 2017]

Contents

Introduction

1. General guidelines

Appendices

A. Language-specific community conventions and/or system related practices

B. OCLC non-Latin script resources

C. PCC documentation upon which this document is based

D. Procedure for updating this document

E. Background/History

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

Introduction

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 (non-Latin (vernacular) script and romanization)^[2]. These guidelines are applicable to operating environments where current RDA-MARC cataloging is used. However, references are made to OCLC, given its importance as an operating environment and as a union catalog for most PCC libraries.

These guidelines apply to all scripts/languages currently supported in OCLC. These include the MARC-8 repertoire of UTF-8 (Chinese, Japanese, Korean, Hebrew, Yiddish, Arabic, Persian, Greek, Cyrillic (Slavic); see <http://www.loc.gov/marc/specifications/specchartables.html>) as well as all Unicode scripts outside of the MARC-8 character set. All these scripts may be used in BIBCO records, but only those scripts included in the MARC-8 repertoire may currently be used in NACO, SACO and CONSER records. As more scripts/languages become supported in the future, this document will be updated as necessary.

Although the decision to include data in non-Latin form in any PCC bibliographic record is strictly optional, when that option is exercised, it should be done according to these guidelines. Catalogers adding data in non-Latin form to records are encouraged to consider the future use, international implications of their records, and to include as much non-Latin script data as necessary to facilitate the identification and location of often scarce material.

Definition of Terms used

Non-Latin script: text in its original writing system not based on the Latin alphabet.

Transliteration: systematic conversion of text from one script to another.

Romanization: transliteration of text into Latin script.

Non-standard romanization: transliteration of text into Latin script using a system other than the ALA-LC Romanization Tables.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

Macros and automatic transliteration

Although the use of macros or other software can greatly facilitate the romanization process, they cannot be depended on to produce correct results in every case, and their results should always be reviewed by catalogers with specialist knowledge of the script and language in question (see also appendix A.).

Examples illustrate RDA and PCC practice. However, RDA instructions and PCC practices may not apply to parallel non-Latin fields in all cases.

1. General guidelines

Romanization of non-Latin scripts must follow the ALA-LC Romanization Tables (<http://hdl.loc.gov/loc.standards/docs.roman>).

All authorized access points (AAPs) must be established in Latin script according to current PCC practice regardless of the presence or absence of non-Latin script elsewhere in the record.

Parallel non-Latin data are encoded in MARC 880 fields in bibliographic records. However, in the OCLC operating environment, parallel fields display as the same MARC tags as their linked Latin equivalent (although indicators may vary). The examples used in these guidelines follow the OCLC convention. Note that unpaired non-Latin fields may be used for fields where pairing is not required by these guidelines. 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.

If non-Latin data is entered in a bibliographic record, in almost all cases, both Latin and parallel non-Latin fields must be created for those fields, parts of fields, or subfields listed in the "Mandatory (if applicable) (MA)" column of the table below. These fields are considered mandatory (if applicable), as they are descriptive, and reflect data transcribed from the resource. Controlled AAPs must be entered in their authorized Latin form, and may also be entered in parallel non-Latin fields. Regarding required or optional elements for specific MARC

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

21 formats, refer to the particular MARC and PCC standards. The “optional” column in the table below is not meant to be exhaustive, but lists commonly used fields.

Mandatory (if applicable) (MA)	Optional
245 (Title/statement of responsibility)	100-111 (Creator/Author authorized access point)
246 (Varying form of title)	130/240 (Work/Expression authorized access point)
247 (Former title)	362 (Dates of publication and/or sequential designation)
250 (Edition statement)	5XX (Notes)
264 (Publication, distribution, etc.)	600-630, 651 (Subject access fields)
490 (Series statements)	700-730 (Additional authorized access points)
740 (Related work/ expression/manifestation)	76X-78X (Related work, expression, manifestation)
	8XX (Authorized access point for series)

1.1. 008 field

Code the fixed field Modified Record Code (008/38) # (blank). If the resource described is in more than one non-Latin script and some of the non-Latin text cannot be entered in its original script because the record is a CONSER record and that script is outside the MARC-8 repertoire (e.g. Tibetan), the record should still be coded with 008/38 # (blank) as long as it also contains at least one script (e.g. Chinese) that has been entered according to these guidelines.

1.2. Fields supplied by ISSN centers: 210 (abbreviated title), 222 (key title)

Do not attempt to provide non-Latin data for these fields.

1.3. Descriptive fields (MA) (245, 246, 247, 250, 264, 490, 740)

When catalogers choose to include non-Latin data in their records, for these descriptive elements parallel non-Latin script data must be given that is transcribed from the resource itself or which is otherwise prescribed by the cataloging rules. Abridgement of the title proper or other title information is according to the cataloging rules. In any such abridgement, that of the non-Latin data must parallel that of the Latin data.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

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-letii u` Nikolai`a` Nikolaevicha Lisovogo / #c otvetstvennyĭ redaktor, A.V. Nazarenko.

245 03 #a די יידישע דראמע פון 20סטן יארהונדערט.

245 03 #a Di Yidishe drame fun 20sṭn yorhundert.

245 10 #a هالة سرحان / #c وقالت شهرزاد : #b أكتب لكم من الحرامك

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

245 00 #a Траектория поэта / #c редактор: Надежда Жильцова ; составители: Надежда Жильцова [and 4 others].

245 00 #a Traektorii`a` poëta / #c redaktor: Nadezhda Zhil't`s`ova ; sostaviteli: Nadezhda Zhil't`s`ova [and 4 others].

[Note: cataloger-supplied information is recorded in the language preferred by the cataloging agency, English for the PCC.]

245 10 #a Грузинские документы IX-XV вв.

245 10 #a Gruzinskie dokumenty IX-XV vv.

246 13 #a Грузинские документы девятого-пятнадцатого вв.

246 13 #a Gruzinskie dokumenty devi`a`togo-pi`a`tnad`t`s`atogo vv.

264 #1 #a تهران : #b انتشارات بهمن برنا، #c 1385 [2006]

264 #1 #a Tihrān : #b Intishārāt-i Bahman Burnā, #c 1385 [2006]

264 #1 #a ירושלים : #b [publisher not identified], #c 2014.

264 #1 #a Yerushalayim : #b [publisher not identified], #c 2014.

[Date in Arabic numerals on resource]

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 14 العدد

490 0# †a Silsilat al-nuṣūṣ al-qānūniyah ; †v al-‘adad 14

1.3.1. Mixed script descriptive fields:

For any descriptive field that includes both Latin and romanized non-Latin data, supply a parallel original-script field that records Latin and non-Latin data in their original form as found on the piece. The 245 second (nonfiling characters) indicator may differ in paired fields, because romanization does not always include the same number of characters as the original script it represents. Some of the languages affected include Hebrew, Arabic, and Greek.

Examples:

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-qaḍā’iyah = †b Le journal judiciaire.

[Example of 245 different second (nonfiling) indicator]

245 00 †a Weekly sun ビジネスニュース.

245 00 †a Weekly sun bijinesu nyūsu.

[Title in Japanese with English as part of phrase (i.e., not a parallel title)]

264 #1 †a Москва : †b MediaPress

264 #1 †a Moskva : †b MediaPress

[Publisher name appears in Latin script on the piece]

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

1.3.1.1. Character sets not supported for CONSER records

If a field including non-Latin scripts in MARC-8 (e.g., Chinese) also includes text in a script that is not supported in MARC-8 (e.g., Tibetan), enter the unsupported text in the same romanized form used in the Latin field.

Example:

245 00 #a Bod kyi sgog pa = #b 西藏大蒜 / #c 'Jigs-med kyis rtsom sgrig byas ; Tshe-nor dan Dpal-'dzoms kyis bsgyur.

245 00 #a Bod kyi sgog pa = #b Xizang da suan / #c 'Jigs-med kyis rtsom sgrig byas ; Tshe-nor dan Dpal-'dzoms kyis bsgyur.

[Parallel field added in order to display original-script Chinese; Tibetan script is outside MARC-8 so the Tibetan text is entered in romanized form in both fields]

While OCLC now supports all Unicode scripts, remember that only those scripts included in the MARC-8 repertoire may currently be used in CONSER records. See the Introduction to this document for more details.

Alternatively, Numeric Character References (NCRs) may be used to represent unsupported scripts, but should be used only in parallel non-Latin fields.

1.4. Notes (optional) (5XX, 362)

In constructing notes, give non-Latin data that appear on the resource itself. Do not translate introductory phrases or other cataloger-created English language data into other languages and scripts.

Examples:

362 1# #a Began with Август 2004.

362 1# #a Began with Avgust 2004.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

362 1 \$a Began with: 第1卷 (平成2年[1990]); ceased with: 第15卷 (平成16年[2004]).

362 1 \$a Began with: dai 1-kan (Heisei 2-nen [1990]); ceased with: dai 15-kan (Heisei 16-nen [2004]).

500 ## †a At head of title: Императорское православное палестинское общество.
Научный совет РАН "Роль религий в истории." Паломнический центр
Московского патриархата.

500 ## †a At head of title: Imperatorskoe pravoslavnoe palestinskoe obshchestvo.
Nauchnyĭ soviet RAN "Rol' religii v istorii." Palomnicheskiĭ t's`entr Moskovskogo
patriarkhata.

500 ## †a Editor: 青木昌彦.

500 ## †a Editor: Aoki Masahiko.

505 0# †a כרך א. חסידים ובעלי מוסר -- כרך ב. משכילים, מתנגדים ורבנים.

505 0# †a Kerekh 1. Ḥasidim u-Va'ale musar -- kerekh 2. Maškilim, mitnagdim ve-rabanim.

505 0# †a Ы́лар саров гнатыр = Поднялась ты гору = Yelar sarov gnatsir -- Ալ էրեսդ ալա է
= Красноликая красавица = Al yeresd ala e -- Կանցի պար = Женский танец =
Woman dance.

505 0# †a Elar sarov gnats'ir = Podni`a`las' ty goru = Yelar sarov gnatsir -- Al eresd ala
ē = Krasnolikai`a` krasavit`s`a = Al yeresd ala e -- Kanats'i par = Zhenskii tanet`s` =
Woman dance.

*[Contents note for a resource with parallel titles in different languages and scripts,
including some "found" romanizations.]*

550 ## †a Official organ of: 한국 관세 연구소.

550 ## †a Official organ of: Han'guk Kwanse Yŏn'guso.

580 ## †a Split into: دراسات قانونية , and: دراسات انسانية.

580 ## †a Split into: Dirāsāt qānūniyah, and: Dirāsāt insāniyah.

1.5. Authorized access points (AAP) (optional) (1XX, 240, 600-630, 651, 700-730, 8XX)

Authorized Access Points are always entered in their authorized, Latin-script form. Non-Latin data may be supplied in parallel fields in the original language/script. When PCC catalogers do supply parallel fields for AAPs, in bibliographic records, they are strongly encouraged to replicate this intellectual work as non-Latin variants in authority records as part of NACO.

This section has two parts: **1.5.1. AAPs: Standard practice** delineates practice for catalogers who are applying a strictly parallel approach to adding non-Latin data for AAPs; **1.5.2. AAPs: Optional practice** is a more flexible approach allowing for enhanced discoverability and catalog display for a particular language or script when needed. In selecting either option, catalogers need to apply judgment based on a thorough knowledge of the script/language in which they are working and the practices of the relevant community.

1.5.1. Authorized access points: Standard practice

Non-Latin data are supplied only for AAPs and parts of them established according to standard, systematic romanization (i.e., according to the ALA-LC Romanization Tables).

- Do not supply non-Latin data for AAPs or parts of them established in non-standard romanization or in a conventional Latin-script form.
- Do not supply non-Latin data for cataloger-created qualifiers (use the established, Latin-script qualifiers).
- Provide non-Latin script additions (e.g., fuller form of name) to personal names that are established according to usage in standard, systematic romanization.
- Do not supply non-Latin data for topical subject or genre AAPs (MARC tag 650, 655) including geographic subdivisions (#z) used in 650s.
- For 600-630, and 651 non-Latin data fields: code the second indicator “4” (Source not specified). Please see Subject Headings Manual, H 181 Arabic/Persian Subject Headings and H 182 Chinese/Japanese/Korean Subject Headings for more details for those languages.
- English relationship designators in parallel original script AAPs are optional.

Examples:

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

100 1# †а Широкорад, А. Б. †q (Александр Борисович)

100 1# †а Shirokorad, A. B. †q (Aleksandr Borisovich), †e author.

[Qualifier is based on standard romanization and not cataloger-created; English relationship designator in parallel original script AAP is optional]

710 2# †а Институт социальных и общинных работников (Minsk, Belarus), †e issuing body.

710 2# †а Institut sot͡s͡ialʹnykh i obshchinnykh rabotnikov (Minsk, Belarus), †e issuing body.

111 2# †а 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# †а Dang dai yu yan xue li lun he Han yu yan jiu guo ji xue shu bao gao hui †d (2006 : †c Beijing, China), †j author.

100 1# †а Dostoyevsky, Fyodor, †d 1821-1881, †e author.

240 10 †а Преступление и наказание. †l English

240 10 †а Prestuplenie i nakazanie. †l English

245 10 †а Crime and punishment

[Preferred title is based on standard romanization; AAP for the author is not]

130 0# †а Преступление и наказание (Motion picture)

130 0# †а Prestuplenie i nakazanie (Motion picture)

245 10 †а Преступление и наказание

245 10 †а Prestuplenie i nakazanie

700 1# †i Motion picture adaptation of (work): †а Dostoyevsky, Fyodor, †d 1821-1881. †t Преступление и наказание.

700 1# †i Motion picture adaptation of (work): †а Dostoyevsky, Fyodor, †d 1821-1881. †t Prestuplenie i nakazanie.

[Russian language film]

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

- 100 0# †a حافظ, †d active 14th century.
- 100 0# †a Ḥāfīz, †d active 14th century, †e author.
- 245 14 †a The Divan of Hafez : †b a bilingual text, Persian-English
- 700 02 †i Container of (work): †a حافظ, †d active 14th century. †t دیوان.
- 700 02 †i Container of (work): †a Ḥāfīz, †d active 14th century. †t Dīvān.
- 700 02 †i Container of (expression): †a حافظ, †d active 14th century. †t دیوان. †l English.
- 700 02 †i Container of (expression): †a Ḥāfīz, †d active 14th century. †t Dīvān. †l English.
[English translation with original Persian]
- 610 24 †a 中国共产党 †x History †y 20th century.
- 610 20 †a Zhongguo gong chan dang †x History †y 20th century.
- 700 1# †a 최 복규, †d 1969-
- 700 1# †a Ch'oe, Pok-kyu, †d 1969- †e author.
- 710 2# †a 中央日報 (Firm), †e issuing body.
- 710 2# †a Zhong yang ri bao (Firm), †e issuing body.
- 710 2# †a 神戸海洋气象台, †e issuing body.
- 710 2# †a Kōbe Kaiyō Kishōdai, †e issuing body.
- 710 2# †a شورای پژوهشهای علمی کشور
- 710 2# †a Shūrā-yi Pizhūhishā-yi 'Ilmī-i Kishvar, †e issuing body.
- 830 #0 †a Жизнь замечательных людей (Molodai`a` gvardii`a` (Firm)) ; †v вып. 1178.
- 830 #0 †a Zhizn' zamechatel'nykh li`u`deĭ (Molodai`a` gvardii`a` (Firm)) ; †v vyp. 1178.
- 830 #0 †a (Sāzmān-i Muṭāla'ah va Tadvīn-i Kutub-i 'Ulūm-i Insānī-i Dānishgāh'hā (Iran))
زبان و ادبیات انگلیسی
- 830 #0 †a Zabān va Adabīyāt-i Ingilīsī (Sāzmān-i Muṭāla'ah va Tadvīn-i Kutub-i 'Ulūm-i
Insānī-i Dānishgāh'hā (Iran))

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

Examples of AAPs or parts of AAPs established in non-standard romanization or in a conventional Latin-script form, for which parallel non-Latin data are not normally entered (see also 1.5.2. for optional practice):

100 0# #a John, #c of Kronstadt, Saint, #d 1829-1909, #e author.

100 1# #a Shidmehr, Nilofar, #e author.

600 10 #a Chiang, Kai-shek, #d 1887-1975 #v Diaries.

700 1# #a Cheung, Leslie, #d 1956-2003, #e actor.

710 1# #a Saint Petersburg (Russia). #b Законодательное собрание.

710 1# #a Saint Petersburg (Russia). #b Zakonodatel'noe sobranie, #e issuing body.

[Subordinate AAP is in standard, systematic romanization]

730 0# #a Ali Baba (Folk tale)

1.5.2. Authorized access points (AAP): optional practice

Several optional exceptions to the above standard practice (1.5.1.) are described below. These are provided to allow the inclusion of non-Latin scripts to parallel AAPs that are not the result of standard, systematic romanization (1.5.2.1.), to mitigate difficulties that often result from mixed-directional script (e.g., Arabic, Hebrew, Persian, etc.) AAPs (1.5.2.3.), and to strongly encourage the inclusion of non-Latin script when it is considered helpful to the catalog user. The practices are described separately, as it is intended that they should not be used by individual language/script cataloging communities that have other guidelines that would contradict these options. In deciding to apply a particular option, consider whether it would better assist the user in searching for and locating the resource being described.

1.5.2.1. Parallel fields for AAPs not established in standard romanization

Non-Latin data may be supplied in parallel fields for AAPs established in non-standard romanization or in a conventional Latin-script form. The parallel non-Latin field for an AAP must be in the language/script most closely associated with the AAP or which is most relevant for the AAP (not necessarily the language/script found on the piece), and the form entered can be derived from the resource itself or if necessary from a standard reference source in the language/script of the AAP. When possible, prefer a non-Latin form that corresponds most closely to the authorized, established Latin form in the AAP.

Examples:

100 1# #a Достоевский, Федор, #d 1821-1881, #e author.

100 1# #a Dostoyevsky, Fyodor, #d 1821-1881, #e author.

100 0# #a Иоанн, #c Кронштадский, Святой, #d 1829-1909.

100 0# #a John, #c of Kronstadt, Saint, #d 1829-1909, #e author.

100 1# #a شیدمهر، نیلوفر.

100 1# #a Shidmehr, Nilofar, #e author.

651 #4 #a 臺東縣 (Taiwan)

651 #0 #a Taitung County (Taiwan)

[Note: does not comply with SHM H 182.]

600 14 #a 蔣介石, #d 1887-1975 #v Diaries.

600 10 #a Chiang, Kai-shek, #d 1887-1975 #v Diaries.

[Note: does not comply with SHM H 182.]

710 1# #a Санкт-Петербург (Russia). #b Законодательное собрание.

710 1# #a Saint Petersburg (Russia). #b Zakonodatel'noe sobranie, #e issuing body.

[Qualifier "Russia" remains in Latin form in this example; however, see also 1.5.2.2.]

1.5.2.2. Entering cataloger-created qualifiers in non-Latin script

In non-Latin parallel fields, cataloger-created qualifiers may be entered in a non-Latin form. The qualifiers must be in the language/script most closely associated with the AAP or which is most relevant for the AAP (not necessarily the language/script found on the piece), and the form entered can be derived from the resource itself or if necessary from a standard reference source in the language/script of the AAP. When possible, prefer a non-Latin form that corresponds most closely to the authorized, established Latin form of the qualifier.

Note: Please see Subject Headings Manual, H 181 Arabic/Persian Subject Headings and H 182 Chinese/Japanese/Korean Subject Headings for non-Latin script practices for 6xx fields in those languages/scripts.

Examples:

710 2# †a Институт социальных и общинных работников (Минск, Беларусь)

710 2# †a Institut sot̄s̄ial'nykh i obshchinnykh rabotnikov (Minsk, Belarus), †e issuing body.

not:

110 2# †a Институт социальных и общинных работников (Мінск, Беларусь)

[AAP is in Russian so non-Latin qualifier must be in Russian, not Belarusian]

710 1# †a Санкт-Петербург (Россия). †b Законодательное собрание.

710 1# †a Saint Petersburg (Russia). †b Zakonodatel'noe sobranie, †e issuing body.

830 #0 †a رواية (بيروت, لبنان)

830 #0 †a Riwāyah (Beirut, Lebanon)

830 #0 †a Монограмма (Москва, Россия)

830 #0 †a Monogramma (Moscow, Russia)

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

830 #0 #a زبان و ادبیات انگلیسی (سازمان مطالعه و تدوین کتب علوم انسانی دانشگاهها (ایران))

830 #0 #a Zabān va Adabīyāt-i Ingilīsī (Sāzmān-i Muṭāla‘ah va Tadvīn-i Kutub-i ‘Ulūm-i Insānī-i Dānishgāh’hā (Iran))

1.5.2.3. Omitting dates and cataloger-created qualifiers (right-to-left scripts only)

For right-to-left scripts only, dates and cataloger-created qualifiers may be omitted from non-Latin parallel fields if it is thought the display may be confusing for the catalog user.

Examples:

100 1# #a صادق زاده، عین الدین

100 1# #a Ṣādiq’zādah, ‘Ayn al-Dīn, #d 1965- #e author.

710 2# #a בית המדרש יד־סופר

710 2# #a Bet ha-midrash Yad-Sofer (Bene Berak, Israel)

710 2# #a מכון המקדש

710 2# #a Temple Institute (Jerusalem)

1.6. Linking entries (optional) (containing authorized access points and identifiers for the related resource (76X-78X))

Give parallel non-Latin script fields only when: a) the linked record is a record containing non-Latin script data (and the non-Latin script data are, therefore, readily available); and, b) the record on which the link is being recorded has other non-Latin script data.

Examples:

785 00 #a Moscow (Russia). Мэрия. #t Вестник Мэра и Правительства Москвы #w
(DLC) 2007222946 #w (OCoLC)51887397

785 00 #a Moscow (Russia). Мэри́я. #t Vestnik Méra i Pravitel'stva Moskvyy #w (DLC)
2007222946 #w (OCoLC)51887397

785 00 #a Majallat Jāmi'at 'Adan lil-'ulūm al-ijtimā'īyah wa-al-insānīyah #w (DLC)
2003546128 #w (OCoLC)60453100

[In the above example, the related record has non-Latin script data but the record on which the link is being added is romanized-only.]

Appendices

A. Language-specific community conventions and/or system related practices

This section contains guidelines and examples that are unique to a particular language and script. This section should be expanded appropriately as new rules for a particular language are needed. Also, as the available character set is expanded, new sections containing guidelines for the added languages/scripts can be included.

A.1. Arabic/Persian

A.1.2. Letters converted incorrectly when using a macro to create script:

In Persian, when using the OCLC macro to create script, the Latin letter h (corresponding to Persian ه) is omitted in conversion when it appears adjacent to the miagkii znak in romanization and manual intervention is required to produce the correct spelling of the word in original script.

Example (after running macro):

245 00 #a Nāmah'hāyī az Īrān.

245 00 #a نامه ایی از ایران

Example (after correction):

245 00 #a Nāmah'hāyī az Īrān.

245 00 #a نامه‌هایی از ایران

Similarly, Latin ā (corresponding to Persian آ) in the middle of words is rendered as I and has to be corrected:

Example (after running macro):

girdāvarī

گرداوری

Example (after correction):

girdāvarī

گردآوری

A.2. Chinese/Japanese/Korean

A.2.1. For personal names given in a parallel non-Latin script field, do not follow a surname with a comma.

A.2.2. Input spaces between lexical units in non-Latin fields that consist of Korean hangul, or hangul and ideographs, or any ideographs that are romanized as Korean language words, so that the non-Latin data mirror the spacing in the parallel romanized fields. Do not input spaces between characters in non-Latin data that are in Japanese or Chinese scripts representing Japanese or Chinese words. For the center dot, input the correct diacritic [Unicode U+30FB], and you may use the input code from the OCLC CJK Dictionary.

Examples:

245 00 #a 壬辰戦争 : #b 16世紀日・朝・中の国際戦争 = 임진 왜란, 동 아시아 삼국 전쟁 / #c 鄭杜熙,李璟珣編著 ; 金文子監訳 ; 小幡倫裕訳. [Spacing added to Korean language to match romanization, spaces not added to Japanese language]

245 00 #a Jinshin Sensō : #b 16-seiki Nitchōchū no kokusai sensō = Imjin Waeran, Tong Asia samguk chōnjaeng / #c Tei Toki, Ri Eishun hencho ; Kin Bunshi kan'yaku ; Obata Michihiro yaku.

700 1# #a 鄭杜熙.

700 1# #a Chōng, Tu-hūi.

700 1# #a 李璟珣.

700 1# #a Yi, Kyōng-sun.

245 10 #a 輸銀年鑑. [*Korean language represented in Chinese characters, with spacing between words*]

245 10 #a Suŭn yŏn'gam. [*Korean romanization*]

245 10 #a 農業団体史・農民運動史 / #c 戦後日本の食料・農業・農村編集委員会編.

245 10 #a Nōgyō dantaishi, nōmin undōshi / #c Sengo Nihon no Shokuryō Nōgyō Nōson Henshū linkai hen.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

245 00 #a통신사 · 왜관 과 한일 관계 / #c 한일 관계사 연구 논집 편찬 위원회 편.

245 00 #a T'ongsinsa, Waegwan kwa Han-Il kwan'gye / #c Han-Il Kwan'gyesa Yǒngu Nonjip P'yǒnch'an Wiwǒnhoe p'yǒn.

A.2.3. Optionally, provide an unpaired variant title (246) with hangul characters for Korean language publications if the title proper consists of hanja (Chinese ideographs) or a mix of hanja and hangul.

Example:

245 10 #a 地方 工業 의 特性 과 育成 政策 / #c 金 鍾基, 黃 一清.

245 10 #a Chibang kongöp ūi t'üksǒng kwa yuksǒng chǒngch'aek / #c Kim Chong-gi, Hwang Il-ch'ǒng.

246 3# #a 지방 공업 의 특성과 육성 정책

A.3. Cyrillic

A.3.1. Characters no longer omitted in romanization

ALA-LC Romanization Tables previously called for omitting a terminal hard sign “ъ” in Russian (pre-revolutionary (1917) script). For Bulgarian, the tables also previously instructed the cataloger to disregard the soft sign “ь” and to disregard the character “ъ” when it appears at the end of a word. Thus, many existing records are still transliterated this way, but the present, revised transliteration tables do not permit omission of these characters in the transliteration.

Examples:

Russian current transliteration:

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

245 10 #a Vı̄e`sti iz" ěkspedit̄s`ti P.K. Kozlova : #b iz" pisem" k" sekretari`u` Obshchestva : s" prilozheniēm" karty.

Bulgarian current transliteration:

245 10 †a Руска книжовна речъ : †b въ образци.

245 10 †a Ruska knizhovna rech' : †b v" obrazt's`i.

A.4. Greek

A.4.1. It is not required to enter the diacritics normally used with Greek script in PCC records. However, these diacritics may be input using the same MARC-8 combining diacritics that are used with Latin-script records. For text printed in capital letters without diacritics, the appropriate diacritics may be supplied. Catalogers may add diacritics to existing records containing Greek script without diacritics. Once input, diacritics that are correct should not be removed from master records.

Example:

245 00 †a Οψεις του αρχαιου Ελληνικου κοσμου : †b (αν)επικαιρες αρχαιογνωστικες συμβουλες.

245 00 †a Opseis tou archaiou Hellēnikou kosmou : †b (an)epikaires archaiognōstikes symvoules.

or:

245 00 †a Όψεις του αρχαίου Ελληνικού κόσμου : †b (αν)επίκαιρες αρχαιογνωστικές συμβουλές.

245 00 †a Opseis tou archaiou Hellēnikou kosmou : †b (an)epikaires archaiognōstikes symvoules.

A.5 Hebrew/Yiddish

A.5.1. When producing Hebrew script, such as that used with OCLC, punctuation problems may accrue:

For example, the OCLC/Joel Hahn macro will introduce hyphens and nekudot, or grammar points that are not transcribed. The vav and yud are usually missing.

Example (after running macro):

245 10 #a הבטם הסטורים, סוציולוגים ותרבתים #b : מחקרים ביהדות קוקז, גאורגיה ובכרה #c / סרג ונשטן
בערכת גולדה ילעזר, ראבן אנוך, סרג ונשטן.

245 10 #a Meḥqarim be-Yahadut Qavqaz, Ge'orgyah ve-Bukharah : #b hebeṭim
hiṣṭoriyim, sotsyologiyim ve-tarbutiyim / #c ba-'arikhat Goldah Eli'ezer, Re'uven
Enokh, Sergei Vinshtain.

Example (after correction):

245 10 #a היבטים היסטוריים, סוציולוגיים #b : מחקרים ביהדות קוקז, גאורגיה ובכרה #c / ותרבותיים
בעריכת גולדה אליעזר, ראבן אנוך, סרגיי וינשטיין #c / ותרבותיים.

245 10 #a Meḥqarim be-Yahadut Qavqaz, Ge'orgyah ve-Bukharah : #b hebeṭim
hiṣṭoriyim, sotsyologiyim ve-tarbutiyim / #c ba-'arikhat Goldah Eli'ezer, Re'uven
Enokh, Sergei Vinshtain.

A.5.2. Names written in the Hebrew alphabet follow the alternative according to LC-PCC PS 9.2.2.5.3 when more than one form of the name is found. The preferred form is the romanized form on the resource. Consequently, many AAPs are established according to a found romanization. The parallel non-Latin field for an AAP is the one found on the item being cataloged.

Example:

100 1# #a לוריא, תקוה וינשטוק

100 1# #a Louria, Tikwah Wenstock, #e author.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

A.5.3. Provide an unpaired variant title (246) for a variant orthographic representation of the title proper if that variant appears elsewhere in the item being cataloged.

Example:

245 10 #a עניין-היהודים

245 10 #a 'Inyan-ha-Yehudim

246 13 #a עניין-היהודים

A.5.4. If the date of publication is represented only in Hebrew letters, apply the second alternative in RDA 1.8.2 and record the date in script in the non-Latin field and supply the equivalent to the found numerals in the romanized field.

Example:

264 #1 #a תשי"ד , החברה לחקירות ארץ ישראל ועתיקותיה #b : ירושלים #a

264 #1 #a Yerushalayim : #b ha-Ḥevrah la-ḥaḳirat Erets-Yiśra'el ve-'atīḳoteha, #c 714
[1953 or 1954]

[Also shows optional addition of Gregorian calendar date.]

A.6. Armenian

A.6.1. When producing Armenian script using a macro, such as that used with OCLC, punctuation problems may accrue:

For example, the OCLC/Joel Hahn macro will change a comma to a “bootnshan” (grave, spacing), and must be switched back to a comma in the Armenian script field in certain cases, depending on the Armenian grammar.

Example (after running macro):

245 00 †a Hakob Ōshakan / †c kazmets' ev khmbagrets', Lilit' Galastyaně.

245 00 †a Հակոբ Օշական / †Ը կազմեց եւ խմբագրեց՝ Լիլիթ Գալաստյանը.

Example (after correction):

245 00 †a Hakob Ōshakan / †c kazmets' ev khmbagrets', Lilit' Galastyaně.

245 00 †a Հակոբ Օշական / †Ը կազմեց եւ խմբագրեց, Լիլիթ Գալաստյանը.

B. OCLC non-Latin script resources

OCLC Connexion Client Guides: International Cataloging: Use Non-Latin Scripts. Revised: July 2016.

<http://www.oclc.org/content/dam/support/connexion/documentation/client/international/internationalcataloging.pdf>

Connexion Client Problems and Troubleshooting: Transliteration, Diacritics, Non-Latin Scripts.

http://www.oclc.org/support/services/connexion/client_known_problems.en.html/#transliteration

OCLC WorldShare Metadata: Record Manager User Guide.

http://www.oclc.org/content/dam/support/worldshare-metadata/Record_Manager_User_Guide.pdf

- Transliterating Latin Text to Non-Latin Text, pages 159-160.
- Unlink Non-Latin Fields, pages 281-282.
- Link Non-Latin Fields, page 470.

PCC Guidelines for Creating Bibliographic Records in Multiple Character Sets

Revised for RDA, June 28, 2016; last revised: September 7, 2017

C. 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-Latin script for Arabic serials [2003]. Appendix O. Creating records with data in non-Latin script for Chinese, Japanese, and Korean serials [updated Sept. 2008].

D. Procedure for updating this document

The environment for including non-Latin data in bibliographic records is developing and changing quickly, and more complexity of instruction for some languages/scripts may be needed, therefore it is expected that this document will need to be periodically revised. Please contact the current chair of the PCC Standing Committee on Standards with requests for revision proposals.

E. Background/History

Based upon recommendations made at the PCC Operations Committee meetings in 2008, a task force of language experts was charged in Feb. 2009 with creating these guidelines (<http://www.loc.gov/aba/pcc/scs/documents/PCCNonLatinTFDraftCharge.pdf>). A preliminary report and guidelines was submitted to the PCC Standing Committee on Standards (SCS) and to PCC discussion lists for feedback in Sept. 2009 (<http://www.loc.gov/aba/pcc/scs/documents/Non-Latin-TF-Prelim-Report-Guidelines.pdf>). The guidelines were first published in May 2010; this version has been revised to reflect RDA.

[1] Standards for Romanization: ALA-LC Romanization Tables: Transliteration Schemes for Non-Latin Scripts (1997 ed.), Washington: Library of Congress, Cataloging Distribution Service, 1997. <http://hdl.loc.gov/loc.standards/docs.roman> Specific scripts may also have further documentation regarding Romanization.

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