

## Thai cataloguing

In November 2005 the National Library implemented an upgrade to its national system that included Unicode support. This has brought many benefits to the Australian library community including the opportunity to provide a wider coverage of non-Roman script cataloguing. (*See: Appendix 1. Supporting non-Roman script cataloguing in Libraries Australia*)

Australia is a multicultural society and libraries collect many community language collections. Hoping to support these collections, some local library systems are already running on Unicode and more are in the process of developing Unicode-based systems. Libraries Australia would like to take advantage of the Unicode implementation at the national and local level and explore the possible expansion of bibliographic description and access to languages other than CJK.

Thai script is the first of these because of the presence of a substantial number of bibliographic records in the National Bibliographic Database<sup>1</sup> and the existence of Thai collections in 18 local libraries. We have recently completed testing Thai script cataloguing in Libraries Australia. The functionality that was tested included record creation, record display, searching and data exchange. The test results are satisfactory and encouraging and we are confident that the support of Thai script is adequate in Libraries Australia and of a similar level to that available in local Unicode systems.

We will continue expanding our support for other non-Roman script cataloguing. There are some limitations in the current Unicode systems. (*See: Appendix 2. Some problems that are caused by Unicode Limitations*) but the aim is to learn from the non-Roman script implementations of other libraries and approach the expansion into non-Roman script cataloguing with patience and proceed in gradual phases. (*See: Appendix 3. Non-Roman script cataloguing in other libraries*)

The following advice pertains to Thai script support; script options for the next stage include Arabic, Hebrew, Tamil, Cyrillic, Greek and other scripts.

---

<sup>1</sup> 36,672 bibliographic records

## **1. Using the Libraries Australia Cataloguing Client only**

You can only perform Thai script cataloguing in the Libraries Australia Cataloguing Client (LACC), but not in the Libraries Australia Search (LAS). Web cataloguing in LAS does not support non-Roman scripts, at least not for now.

NOTE: Validation for web cataloguing needs to allow for input of parallel script.

## **2. Script display and input**

You need to install the required fonts and script input devices on your PC. Here we, like many libraries and system vendors, would like to recommend using the MS Arial Unicode font and the Microsoft Input Method Editor (IME) for the Thai script. The MS Arial font and Thai IME are available as part of the Microsoft Office suite.

The level of support in Windows depends on the operating system version. Windows XP has improved tools for the Unicode environment and we found no obvious problems during our testing. Users who are using older versions such as Windows 98 or 2000 may run into problems as described in *Appendix 2*.

## **3. MARC formats**

If using the LACC to input records, you need to be familiar with the LACC-specific MARC format for Thai records. In brief, this is the use of parallel tags for input, though records are exported using 880 linked fields. You also need know if you are required to use a different MARC format in your local systems to accommodate the Thai records. (See: *Appendix 4. MARC formats for Thai multiscript records*)

Information about CJK script cataloguing can be found in the *Libraries Australia Cataloguing Client Manual: 4.4.4.7. Non Roman scripts specific information*. The information there applies to Thai script.

## **4. Romanized data**

Data is recorded in the appropriate MARC fields using Romanized words. There are many schemes for Thai script Romanization. Libraries in the English speaking world are required to follow the guideline provided by the America Library Association and the Library of Congress. A Thai Romanization table can be found in the ALA-LC Romanization Tables on the Library of Congress website (<http://www.loc.gov/catdir/cpso/Roman.html>).

## **5. Script data**

You need to be familiar with which MARC fields contain the Thai script data. The only mandatory field for non-Roman script data is the 245 field (Title) (See: *Appendix 5. Which MARC fields are required to add Thai script?*)

## 6. Indexing and searching

If Thai script characters are present in indexed fields, they will be searchable in both LACC and LAS. The main issue with searching for Non-Roman script characters is that currently LACC indexes both Roman and non-Roman script words according to word divisions. This has caused problems for searching those scripts that have words written without spaces in between. The system currently treats two, three or more Thai or CJK words that are joined together as one word.

The next LACC release will index both single non-Roman characters and immediately adjacent characters in a field. We believe that this will make word or keyword searching for Thai and CJK scripts more meaningful. Until the next release, for more precision and recall of script, applying truncation after the minimum 3 characters in word or phrase searching will provide more precise search results.

Because of the complexity in indexing and searching Thai and other non-Roman scripts, it is recommended that in LACC or LAS you use the standard Thai Romanization for searching. For more details see: *Libraries Australia Cataloguing Client Manual: 3.2.16. Scripts or Romanised searching*

## 7. Some useful tools

### 7a. Customizing diacritics

Diacritics are a standard feature in ALA-LC Thai Romanization Table. It is common for Romanised Thai words to have single or double diacritics. LACC Special Characters List covers all 10 ALA-LC Thai diacritics. But for easy application, instead of using the Special Characters bar to enter diacritics, you can customise your LACC by adding commonly used diacritics to the toolbar. (I think this should become an appendix) Here are the steps to do that:

- In the LACC editing screen, the Special Characters bar should be on. Under View, ensure that 'Special Characters' is checked.
- On the Script menu, select Learning.
- Click on the required diacritic.
- Click on Script, then Learning again to turn recording mode off.

- You will be asked to name your script. Do not use numbers, punctuation or spaces in your script name. If you must use a space, then substitute an underline instead (eg 'My\_Cedilla' or 'MyCedilla' NOT 'My Cedilla' or 'My:Cedilla').
- A script edit box will appear. Click on the 'x' in the right hand corner to close it.
- The Customise dialogue box will appear (or Options > Customise > Commands) where you can add a shortcut key and/or a button on the toolbar to run your script. The name of your script should appear in Functions under the Commands tab. Click on the function name, keep the mouse button pressed down and drag to the toolbar. (For more information see: *Libraries Australia Cataloguing Client Manual: 2.9.1. Customize*)

You need to know that Unicode has a unique way to store diacritics. In LACC the entry of diacritic must be after the target script they modify. Here are a few examples:

khwāmčhing (khwa|macron|mc|hachek|hing)

kānplǣ (ka|macron|nplǣ|macron|

khrūang (khrū|macron|ang

khō (kho|right cedilla|macron|

### *7b. Automatic Romanization*

Automatic Romanization calls on a built-in program within the cataloguing system which provides standard Romanized words automatically for the original script. Some programs claim that automatic Romanization from the Romanized words into the original script is also possible. This highly sought library system functionality helps to save cataloguing resources; it can also make cataloguing easier for those cataloguers who are more familiar either with the Romanization or the original script.

LACC will soon provide automatic Romanization for Chinese script. We are aware that OCLC can provide the functionality for Arabic and Cyrillic, and will be testing these as the opportunity arises. There are at least two stand-alone

Thai automatic Romanisation programs available and we intend investigate the possibility of providing this functionality for the Thai script in the future.

### *7c. Other databases that have Thai script records*

#### 7c.1. Sources for copy cataloguing

Libraries in the western world that collect Thai materials mostly have Romanized records in their catalogues. OCLC has recently announced support for Thai script in WorldCat. A number of Thai libraries have Thai script in their collections however libraries in Thailand do not strictly follow the MARC21 format. Thai script is used in many regular MARC21 fields even with those fields that usually carry Latin scripts such as 300 (physical description) and 650 (subject topical).

The Thailis Union Catalog (<http://uc.thailis.or.th/> or <http://202.28.18.229/cgi-bin/gw/chameleon?lng=en&skin=cmu>) most closely suits our purpose. The catalogue consists of about 2.9 million bibliographic records, it has an English interface, allows for ISBN searches, allows MARC record direct download, and also uses Unicode encoding for their Thai.

Using mainly the “copy and paste” process in our testing we have found copying Thai script data from the Thailis Union Catalog and pasting them into LACC is a straightforward process.

#### 7c.2 Records download via Z39.50 in LAS

Other databases that have multiscript CJK records are available in the ‘Other Databases’ area in LAS. Users can access these databases via Z39.50 and records can be imported to LAS and LACC in near real time.

Deleted: areain

We have identified two databases in Thailand that provide Z39.50 access. They are the National Library of Thailand and Shinawatra University. Testing will be conducted to ensure connection standards and data quality are acceptable before adding them to the LAS ‘Other Databases’ list.

Should the two remote Thai databases become available in LAS, you will be able to import records into LAS and LACC. As with all records, the cataloguer will have the responsibility of upgrading the imported record to meet Libraries Australia cataloguing standards.

Deleted: cataloguerwill

The following is a sample record from Shinawatra University:

LDR: 00259nam 22001337a 4500  
008: 041022s20042547th 00010 tha  
020: \$a 9743232214 : \$b (pbk.)  
037: \$a CU : \$c 171  
050: 4 \$a JQ1745.A55 \$b 146 2547

- 100: 1 \$a ปราน พิสิฐเศรษฐการ.
- 245: 10 \$a ทักษิณบนเวทีเอเชียและโลก : \$b รวมสุนทรพจน์สำคัญของ พ.ต.ท. ดร.ทักษิณ ชินวัตร / \$c ปราน พิสิฐเศรษฐการ.
- 246: 31 \$a Thaksinomics and Thaksin style strategic thinking : Thaksin's most strategic keynote addresses
- 246: 3 \$a ทักษิณบนเวทีเอเชียและโลก : Thaksinomics III
- 250: \$a พิมพ์ครั้งที่ 1.
- 260: 0 \$a กรุงเทพฯ : \$b มติชน, \$c 2547.
- 300: \$a 303 หน้า ; \$c 21 ซม.
- 440: 0 \$a หนังสือชุดทักษิโนมิกส์และการคิดเชิงกลยุทธ์สไตล์ทักษิณ.
- 536: \$a LIT
- 600: 14 \$a ทักษิณ ชินวัตร. \$c พ.ต.ท., \$d 2492 \$x ทักษะเกี่ยวกับการเมือง.
- 600: 14 \$a ทักษิณ ชินวัตร. \$c พ.ต.ท. \$d 2492 \$x ทักษะเกี่ยวกับเศรษฐกิจ.
- 600: 14 \$a ทักษิณ ชินวัตร. \$c พ.ต.ท. \$d 2492 \$x สุนทรพจน์.
- 650: 4 \$a การบริหารรัฐกิจ \$z ไทย.
- 650: 4 \$a นโยบายของรัฐ \$z ไทย.
- 651: 4 \$a ไทย \$x ความสัมพันธ์กับต่างประเทศ \$z เอเชีย.
- 651: 4 \$a ไทย \$x ความสัมพันธ์ทางเศรษฐกิจกับต่างประเทศ \$z เอเชีย.

*Notes:*

1. Thai script in regular MARC fields
2. Topical subject headings (650) are in Thai script
3. It has some local specific fields

To upgrade the imported record, you are required, for example, to provide parallel fields for some MARC fields, add appropriate subject headings and remove unwanted fields. Please see *Appendix 4* for more about the script cataloguing standards in LACC.

The following is the imported record after the upgrade:

```
003 $0AuCNLKIN
020 ##$a9743232214
040 ##$aANL$beng$dANL
041 0#$atha
043 ##$aa-th---
082 04$a320.9593$222
099 ##$IB$a000026522086
100 0#$601$7ba$aPrān Phisitsētthakān.
100 0#$601$7ia$aปราน พิสิฐเศรษฐการ
245 1#$602$7ba$a@Thaksin bon wēthī □Ēchīa læ lōk :$brūam
```

sunthō nphot samkhan khō ng Phō .Tō .Thō . Doktō Thaksin  
 Chinnawat = Thaksinomics and Thaksin style strategic thinking : Thaksin's  
 most strategic keynote addresses [Prān Phisitsētthakān.  
 245 1#602\$7ia@a@ทักษิณบนเวทีเอเชียและโลก :\$บรวมสุนทรพจน์สำคัญของ  
 พ.ต.ท. ดร. ทักษิณ ชินวัตร /\$cปรารณ พิสิฐเศรษฐการ, รวบรวม  
 246 31\$Thaksinomics and Thaksin style strategic thinking  
 246 30\$Rūam sunthō nphot samkhan khō ng Phō .Tō .Thō . Doktō  
 Thaksin Chinnawat  
 250 ##\$Phim rūam lēm khrang r k.  
 260 ##\$603\$7ba\$Krung Thēp :\$bMatichon,\$c2547 [2004]  
 260 ##\$603\$7ia\$aกรุงเทพฯ : \$bmดิชน, \$c2547  
 300 ##\$a303 p. :\$bill. ;\$c21 cm.  
 440 ##\$604\$7ba\$a@Nangsu chut Thaksinōmik læ kānkhit chōe ng  
 konlayut satai Thaksin  
 440 ##\$604\$7ia\$a@หนังสือชุดทักษิณมิกส์และการคิดเชิงกลยุทธ์สไตล์ทักษิณ  
 490 1#\$aChut wichākān  
 600 00\$605\$7ba\$aThaksin Chinnawat\$d1949-  
 600 00\$605\$7ia\$aทักษิณ ชินวัตร, \$d1949-  
 651 #0\$W000001498948\$aThailand\$xPolitics and government  
 651 #0\$W000001498856\$aThailand\$xEconomic conditions

## 8. Data exchange between Libraries Australia and local systems

MARC21 is the commonly accepted format for data exchange between  
 Libraries Australia and its members. Though the Libraries Australia Record  
 Import Service expects records containing 880 linked fields, the LACC uses  
 parallel fields for input of records. MARC21 records are exported in MARC21  
 using the 880 linked fields.

MARC21 allows the encodings of MARC8 and Unicode for non-Roman script  
 records. Libraries Australia provides the choices of MARC8 or Unicode data  
 format for import/export. Thai script is outside the MARC8 repertoire, and  
 Thai data exchange is only available on Unicode-based systems. Unicode is  
 the only acceptable data format for record export and import.

In case your library system is a Latin or a MARC8 based system, you are,  
 firstly, not able to create Thai script records. Secondly, if you import Thai  
 records from Libraries Australia, no matter what data format you are using,  
 the Thai characters of the records imported to your local system will be  
 unintelligible.

## Appendixes

1. **Supporting non-Roman script cataloguing in the Libraries Australia**
2. **Some problems that are caused by Unicode Limitations**
3. **Non-Roman script cataloguing in other libraries**
4. **MARC formats for Thai multiscrypt records**
5. **Which MARC fields are required to add Thai script?**

### 1. Supporting non-Roman script cataloguing in the Libraries Australia

- Previously the National Library maintained two national databases; Amicus, which contained regular Latin script records, and the National CJK database that contained Latin and CJK multiscrypt script records. Now the two co-exist in Libraries Australia. Some Libraries Australia members who catalogued CJK script records in the CJK database have been making full use of the Unicode implementation and continue to catalogue CJK records in the new national database. Currently Libraries Australia contains 765,000 CJK bibliographic and authority records. In addition, there are 460 serial titles from the Library of Congress that contain Arabic script.
- Users will be able to deliver linguistically accurate displays in ANBD either in Libraries Australia Cataloguing Client (LACC) or in Libraries Australia Search (LAS).
- Users will be able to record in LACC the scripts and characters that appear in the materials being are cataloged.
- Users will be able to search in those scripts and characters in LACC and LAS.
- Users will be able to exchange data between Libraries Australia and its members in a standard international format.
- The new national system is Windows-based and allows the use of the existing fonts and input devices, Microsoft MS Arial Unicode font and the Input Method Editors, for non-Roman script display and input. This is a great achievement as only a few years ago the support for any non-Roman script display and input was considered to be the one of most difficult tasks by system vendors and library administrators.

### 2. Some problems that are caused by Unicode Limitations

The limitations derive from the situation that no existing fonts and character input devices can fully support all Unicode characters.

Thai script has 87 characters, and is well supported in Unicode. The complex display and input problems which are common to CJK users does not usually apply to Thai librarians.

*(a) Script input*

Microsoft's Thai IME or any Unicode-based text editor can be used for Thai character input. We recommend the Thai IME as it provides users with the ability to input Thai scripts on a standard PC keyboard. For more information about IMEs, please see *Libraries Australia Cataloguing Client Manual: Appendix 8. Keyboard layouts and Input Method Editors*.

The performance of Thai IME depends on your PC. If you are using an older operating system, you may find the required Thai character is not available in the IME.

*(b) Fonts*

There are no fonts currently in existence that will display the 65,000 plus characters in Unicode version 4.0. Two fonts are available however which have relatively good coverage of characters in languages commonly encountered. These are "Arial Unicode MS" from Microsoft and "Andale Mono" from Agfa-Monotype.

The Arial Unicode MS font is based on Unicode version 2.1 and includes 38,871 graphic characters. Andale Mono supports all characters in Unicode 3.0; that is, 49,170 graphic characters.

Both Arial Unicode MS and Andale Mono claim that they will undergo the upgrade and the new version will have better quality and more complete coverage of the characters they already covered.

MS Arial Unicode font is widely used in the library community. The Library of Congress uses it in its local system. OCLC and RLG also recommend this font. Arial Unicode MS is one of the available fonts in Microsoft Office. The font was available as a free download from Microsoft but now can only be obtained under licence.

MS Arial font has a long history of supporting what the Library of Congress called the JACKPHY (Japanese, Arabic, Chinese, Korean, Persian, Hebrew and Yiddish) scripts. LC plans in the near future to support Cyrillic, Greek with the MS Arial Unicode font. OCLC depends entirely on MS Arial Unicode to implement Thai and Tamil script cataloguing.

There are at least four problems with the MS Arial Unicode font:

1. Its performance depends on the operating system of the PC;
2. It only supports limited characters;

3. Some scripts, for example, Arabic and Thai rendered in the MS Arial Unicode font are seen as childish and ugly to native speakers; and
4. The Library of Congress has found out that the font cannot support Burmese, Amharic and many other languages and scripts they have in their collection.

Problems 1 and 2 are specifically related to the Thai script. So it may happen when you find the display of the desired characters is not supported, and instead of getting the correct character, you see “?” or some meaningless symbol.

*(c) Possible solution to input and display problems*

c.1 Input

If entering Thai or other non-Roman characters using an IME is not possible, there are many other alternatives. A number can found in the document titled “How to enter Unicode characters in Microsoft Windows” which is available online at: [http://www.fileformat.info/tip/microsoft/enter\\_unicode.htm](http://www.fileformat.info/tip/microsoft/enter_unicode.htm)

We would recommend the use of two options: “ALT + Unicode hexadecimal value” and “Character Map”. The former is mentioned in the above document.

ALT + Unicode hexadecimal value

- Input UCS/Unicode hex values of diacritics and non-Roman script using ALT in combination with the ‘+’ on the number keypad and/or alpha key values e.g. ALT+8eca, ALT+11df.
- For example: Press and hold the ‘ALT’ key. Press the ‘+’ key on your Number Keypad, followed by the Hexadecimal value of the character you wish to enter. Release the ‘ALT’ key.
- Number Lock (Num Lock) should be off.
- The following registry change may be required. We **strongly** recommend that this change be made by your IT support staff.
- To make this change, go to Start > Run > type "regedit" > OK, then navigate to the following path and RHS mouse click > new string value.
- HKEY\_Current\_User/Control Panel/Input Method
- then set EnableHexNumpad to "1".
- If you have to add the value, set the type to “REG\_SZ”.

Character Map

This option is easy to use and does not need IT support. The steps are:

- Go to Start > Programs > Accessories > System tools > Character Map (or Start > Run > type “charmap” > OK)
- The font should be set to Arial Unicode MS
- Click on **Advanced view**
- Enter the Unicode hexadecimal value in the **Go to Unicode** box
- Click on the required character, click on **Select**, then **Copy**.
- In the LACC editing screen, copy the character into your record using Edit>Paste (or the shortcut key Ctrl-V.)

However, in order to use the alternative methods, you need to know the Unicode hexadecimal value or how to find them.

The 87 Unicode Thai characters are in the Unicode range between 0E00 and 0E5B, and you can view them by going to The Unicode character code charts by script (<http://www.unicode.org/charts/>) in the Unicode home page. Select Thai from the South East Asian scripts.

One needs to use search page (<http://www.unicode.org/charts/unihansearch.html>) in the Unicode home page for the complex scripts such as CJK.

### c. 2 Display

If the desired character cannot be displayed, there are also other alternatives which are mentioned in the article of “Assessment of options for handling full Unicode character encodings in MARC21, Part 1 (<http://www.loc.gov/marc/marbi/2004/2004-report01.pdf>)

Deleted: unicode

We would recommend one of the options mentioned in the above article which is to replace the character with “[U+nnnn]” in ASCII. So if you find that the Thai character cannot be displayed, find the Unicode value and replace it with for example [U+0E10]. This keeps the important Unicode value. Some other users who have better display capability will find it easy to put the correct character back.

The “[U+nnnn]” approach would be most useful for CJK librarians.

### **3. Non-Roman script cataloguing in other libraries**

The Library of Congress has a long commitment to non-Roman script cataloguing. Before the conversion of its local system in November 2005 to Unicode, it supported all Latin script cataloguing in its local database and

Deleted: cataloging

added the non-Roman JACKPHY scripts to bibliographic records using specialised cataloguing systems (RLIN and OCLC) for many years. In their new system LC will continue to support the JACKPHY scripts and they also plan in the near future to support Cyrillic and Greek as a possibility. All these scripts are contained in MARC21 repertoire; they can be supported either in MARC8 or Unicode encoding and have full round-trip mapping between the two encodings. LC has yet no plans to go beyond MARC21 repertoire, for example by supporting Thai, Tamil, Hindi, Armenian and other scripts. LC still maintains strong commitment to distribute records to libraries that are using MARC8.

RLIN will probably follow the policy of LC, but OCLC has decided to take a bigger step in their approach concerning non-Roman script cataloguing. In February 2006 they announced that they would continue supporting CJK script and would add the support of Tamil and Thai scripts cataloguing in WorldCat. OCLC will also support Cyrillic soon.

Both LC and OCLC have realised that although Unicode supports a large number of scripts, any Unicode-based system is also facing at least three limitations. Firstly, no single font can support the display of all Unicode characters. Secondly, no single input device can support all Unicode characters. Thirdly, there are “standards compliant” problems for data exchange between Unicode system and non-Unicode system. The first two are system limitations and are universally applied to all Unicode-based systems. Even though there have been some improvement for supporting Unicode display and input in Windows XP, it is not exhaustive. The third problem is related to varying cataloguing policies between libraries.

#### **4. MARC formats for Thai multiscrypt records**

##### **A. Bibliographic records**

Different systems may use different MARC formats for multiscrypt bibliographic records. For example, OCLC uses OCLC MARC, RLIN uses RLIN MARC and LACC uses PICA MARC. But Model A (Vernacular and Transliteration) of Appendix D (Multiscrypt records) of the MARC21 Format for Bibliographic Data Format is commonly adopted by the majority of the libraries. “Model A” details are available in The Library of Congress home page: (<http://www.loc.gov/marc/bibliographic/ecbdmulti.html>).

For details on how PICA MARC supports non-Roman script cataloguing, please see: *Libraries Australia Cataloguing Client Manual: 4.3.3.7. Non Roman scripts specific information.*

The following is an example of a Thai script bibliographic in LACC:

```
003 $0AuCNLKIN
010 ##$a 92929616
020 ##$a9748313433
040 ##$aLC$beng$cLC$dANL
```

050 00\$aMLCSE 94/708 (B)  
 082 04\$a294.3442\$221  
 099 ##\$IB\$a000010897491  
 100 0#\$601\$7ba\$aPhra Thēpwēthī\$Q(Prayut)  
 100 0#\$601\$7ia\$aพระเทพเวที\$Q(ประยูร)  
 245 1#\$602\$7ba\$a@Phō□n talō□t pī, chīwit dī talō□t pai\$cPhra Thēpwēthī (Prayut Payuttō).  
 245 1#\$602\$7ia\$a@พรตลอดปี ชีวิตลลลลไป\$cพระเทพเวที (ประยูร ปยุตโต) .  
 250 ##\$603\$7ba\$aPhim khrang thī 2.  
 250 ##\$603\$7ia\$aพิมพ์ครั้งที่ 2.  
 260 ##\$604\$7ba\$a[Bangkok] :\$bMūnnithi Phutthatham,\$c2535 [1992]  
 260 ##\$604\$7ia\$a[กรุงเทพฯ] :\$bมูลนิธิพุทธธรรม,\$c2535 [1992]  
 300 ##\$a86 p. ;\$c19 cm.  
 500 ##\$aIn Thai.  
 520 ##\$a1991 New Year's lecture, Sirirat Hospital, on blueprint for practical living according to Buddhist principles.  
 650 #0\$aLife\$xReligious aspects\$xBuddhism  
 650 #0\$9351541772Buddhism -- Doctrines  
 650 #0\$935757186XReligious life -- Buddhism  
 710 2#\$605\$7ba\$aMūnnithi Phutthatham  
 710 2#\$605\$7ia\$aมูลนิธิพุทธธรรม

## Notes

1. Leader character position 9: Character encoding scheme is defaulted to "a" which indicates that the record uses Unicode for its character encoding. More information about Leader character position 9 can be found in: (<http://www.loc.gov/marc/bibliographic/ecbdldrd.html>)

2. Some of regular fields (100, 245, 250, 260, and 710) are in pairs. One is in Latin script and one in Thai script. They are linked to each other with the linkage subfield \$6 plus the occurrence number such as 01, 02, 03 and 04.

3. \$7 is the script subfield, "ba" and "ia" are script codes for Latin and Thai respectively.

In your local system, you should use MARC21 for cataloguing Thai script records. LAS stores non-Roman script records in MARC21 and the following shows how the above record will be stored in LAS. This is the correct format for local Unicode libraries to follow.

000 01499cam a2200337 a 4500  
 001 000010897491  
 003 AuCNLKIN  
 005 20060721124556.0  
 008 940418s2535 th 000 0 tha  
 010 \_\_\_ \$a 92929616  
 020 \_\_\_ \$a9748313433  
 040 \_\_\_ \$aLC\$beng\$cLC\$dANL  
 050 00 \$aMLCSE 94/708 (B)  
 082 04 \$a294.3442\$221

100 0\_ \$6880-01\$aPhra The□pwe□thi□\$q(Prayut)  
 245 10 \$6880-02\$aPho□□n talo□□t pi□, chi□wit di□ talo□□t pai /\$cPhra  
 The□pwe□thi□ (Prayut Payutto□).  
 250 \_\_\_ \$6880-03\$aPhim khrang thi□ 2.  
 260 \_\_\_ \$6880-04\$a[Bangkok] :\$bMu□nnithi Phutthatham,\$c2535 [1992]  
 300 \_\_\_ \$a86 p. ;\$c19 cm.  
 500 \_\_\_ \$aln Thai.  
 520 \_\_\_ \$a1991 New Year's lecture, Sirirat Hospital, on blueprint for practical  
 living according to Buddhist principles.  
 650 \_0 \$aLife\$xReligious aspects\$xBuddhism.  
 650 \_0 \$aBuddhism\$xDoctrines.  
 650 \_0 \$aReligious life\$xBuddhism.  
 710 2\_ \$6880-05\$aMu□nnithi Phutthatham  
 850 \_\_\_ \$aANL\$cTHp REL 15/9  
 880 0\_ \$6100-01\$aพระเทพเวที\$q(ประยูรฑ์)  
 880 10 \$6245-02\$aพรตตลอดปี ชีวิตดีตลอดไป /\$cพระเทพเวที (ประยูรฑ์ ปยุตโต) .  
 880 \_\_\_ \$6250-03\$aพิมพ์ครั้งที่ 2.  
 880 \_\_\_ \$6260-04\$a[กรุงเทพฯ] :\$bมูลนิธิพุทธธรรม,\$c2535 [1992]  
 880 2\_ \$6710-05\$aมูลนิธิพุทธธรรม

*Notes:*

1. Leader character position 9 is coded as "a". Most Unicode-based systems have this default value.

2. Some of the regular fields (100, 245, 250, 260, and 710) which contain standard Romanized data are linked to the 880 fields with the linkage subfield 6 plus the occurrence number.

3. The 880 linked fields which contain the Thai script are linked back to the regular fields with the linkage subfield 6, plus the same occurrence number.

4. Information about the 880 fields (Alternate graphic representation) can be found in the following link:

<http://www.loc.gov/marc/bibliographic/ecbdhold.html#mrCb880>

**B. Authority records**

We recommend following the practice that is used for CJK authority records. The practice is established according to MARC21 format for authority data which is to record the preferred form in the 1XX fields, and the script form in the 4XX fields as non-preferred form. For example:

100 1 \$aNiramon Mēthīsuwakun\$d1961-  
 400 1 \$aนิรมล เมธีสุวกุล\$d1961

Currently there are more than 1,400 CJK authority records in Libraries Australia with script data in the 4XX fields.

## 5. Which MARC fields are required to have Thai script?

Even though Unicode-based systems enable users to record Thai script anywhere in a MARC record, there are nevertheless some common rules to be followed, and they can be found in the following two useful documentation:

- Program for Cooperative Cataloging Non-Roman Core Record Task Group Final Report (<http://www.loc.gov/catdir/pcc/archive/jackphy.html>)
- CONSER Editing Guide Appendix O. Creating Records with Data in Non-Roman Script for Chinese, Japanese, and Korean Serials (<http://www.itsmarc.com/crs/edit7839.htm>)

Libraries Australia recommends the following guidelines:

Element	Latin script	Non-Roman scripts
Fixed length data (0XX)	M	M
Main entry (1XX)	MA	MA (2)
Uniform title (240)	MA	O (3)
Description (245-300)	M	M (4)
Series statement (4XX)	MA	O
Notes (5XX)	O	O (5)
Subject access (6XX)	MA	O (6)
Added entries (7XX)	MA	MA (2)

### Footnotes

*M= Mandatory*

*MA= Mandatory if applicable*

*O= Optional*

### Notes

#### 1. Fixed length data

If a non-Roman ISBN is to be entered it must be entered in a subfield z. Please refer to the minimum record standard (<http://www.nla.gov.au/librariesaustralia/network.html>) for the provision of coded data elements in Leader and 008 fields.

#### 2. Main entry and added entries

Name entries may have non-Roman script entries if the Latin entry uses standard ALA/LC Romanization of the script heading. Script headings are required if the Latin entry represents standard ALA/LC systematic Romanization.

Do not supply parallel non-Roman script field for conventional headings that do not conform to standard ALA/LC systematic Romanization.

Give non-Roman script fields in the form found in the publication if it equates to the established form.

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

### 3. Uniform title

LACC users may give parallel fields containing non-Roman script data for variant titles, if the variant titles provide useful access points.

### 4. Description

In the non-Roman script title (245) field, information must be completely transcribed, including title proper, parallel titles, and statement of responsibility as found on the chief source.

### 5. Notes

Non-Roman script general note (500) fields should not be constructed by the cataloguer. Non-Roman script data may be represented in original script. The general note (500) field may be present and entered either in Romanization only or in the non-Roman script only or in parallel fields. Content (505) field should be present and it may be entered either in Romanized form only or in the non-Roman script only or in parallel fields.

### 6. Subject access

Do not supply non-Roman script data for topical subject heading (650) even though there may be a one-to-one equivalent between non-Roman script characters and the established form.