



FKOHA 圖書館管理系統多國語文/多國字體功能之評析

張迺貞¹，蔡育欽²

¹大同大學通識教育中心，台灣

ncchang@ttu.edu.tw

²財團法人國家實驗研究院國家高速網路與計算中心，台灣

thomas@nchc.org.tw

Meeting: 135. UNIMARC

WORLD LIBRARY AND INFORMATION CONGRESS: 75TH IFLA GENERAL CONFERENCE AND COUNCIL

23-27 August 2009, Milan, Italy

<http://www.ifla.org/annual-conference/ifla75/index.htm>

摘要

就圖書館管理系統而言，近年來的趨勢已逐漸重視與發展自由軟體(Free Software)與開放原始碼(Open Source Software, OSS)之系統。圖書館管理系統使用的一些標準，幾乎是普遍應用於世界各地的圖書館，其中，國際機讀編目格式標準 UNIMARC (Universal MARC Format) 有各種不同的衍生機讀編目格式。本文著眼於在開放資源圖書館管理系統 Koha 之下，執行 UNIMARC、CMARC 以及 MARC21 書目資料。比較並評估在代碼資料段欄位以及連接款目段欄位，Koha 支援到何種程度。如果無法執行，需要提供什麼支援。

1. 背景

在過去幾年裡，圖書館自動化系統(the Integrated Library System, ILS)的產業環境發生很大的變化。隨著自由軟體逐漸受到重視，市場趨勢轉向發展開放原始碼(Open Source Software, OSS)圖書館自動化系統。圖書館轉向開放原始碼

系統解決方案，主要是因為相關產品已經發展相當成熟兼具現代化，能滿足圖書館的需求，例如採用通用公共授權(General Public License, GPL)授權的 Koha 系統；開放原始碼系統有利於客製化，能滿足圖書館的特殊需求；開放資源系統的諸多功能來自於使用者社群的回饋，因此社群圖書館都可以使用，不但節省開支還可以從中獲益。此外，圖書館採用開放原始碼系統比起商業系統有更多的支援選擇性(Breeding, 2007)。

圖書館自動化圖書館管理系統比起許多資訊檢索系統要求較高階的文字處理功能，因為它需要考慮到特殊讀音符號。計算機最初是在美國被大量使用，主要使用基本拉丁符號，並不需要考慮到特殊讀音符號；但是當我們檢索資料時，我們必須能夠區分 *eleve* 和 *élève*，也就是說，當我們檢索 *eleve* 時，是否 *élève* 也會被檢索出來。再加上非拉丁字母，更是增加檢索資料時的複雜性。Unicode 就是為了解決這個問題應運而生的。

圖書館管理系統使用的一些標準，幾乎是普遍應用於世界各地的圖書館，其中的一個標準可以說是最重要的，就是 MARC，因為它可以讓所有圖書館分享書目記錄。遺憾的是，它有各種不同的衍生機讀編目格式，其中之一是國際機讀編目格式(Universal MARC Format, UNIMARC)。中國機讀編目格式 (Chinese MARC, CMARC)，是 MARC 次衍生機讀編目格式，因為它是由 UNIMARC 衍生而來，是目前台灣圖書館最廣泛使用的機讀編目格式。美國國會圖書館發展的機讀編目格式 LCMARC，現在稱為 MARC21，是最常見的 MARC 衍生機讀編目格式。幾乎所有的開放原始碼系統都使用 MARC21，但是也有其他的衍生機讀編目格式比 LCMARC/MARC21 具備有後來發展的更新的特色。例如 UNIMARC 就制定出比 MARC21 更高階的分欄細膩深度，並有不同的方法經由軟體執行連接欄位段的資料。

UNIMARC 第 3 版 (2008 年) 將 Unicode 納入，就是考慮到要與國際公認的機讀編目格式結構發展趨勢接軌。2004 年，台灣國家圖書館成立了一個非官方的 Unicode 工作小組，目的是尋求解決中文內碼系統的議題，也就是中文標準交換碼(Chinese Character Code for Information Interchange, CCCII)和大五碼 (Big5) (毛慶禎、許靜芬，2006 年)。這 2 種中文內碼是目前台灣最廣泛使用於各圖書館的內碼系統。

Koha 圖書館管理系統最初是發展用於英語世界的 MARC21 的架構之下。下一章節，我們將討論在 Koha 圖書館管理系統之下，執行 UNIMARC、CMARC 以及 MARC21 書目資料。比較並評估 Koha 支援到何種程度 (例如連接連續性出版品的新舊題名的方法)。如果無法執行，需要提供什麼樣的支援。

Koha 圖書館管理系統最初是 Katipo 通信公司為 Horowhenua 圖書館於 1999 年發展的。2003 年秋季在俄亥俄州 Nelsonville 公共圖書館正式啟用。根據 Breeding (2008 年) 的調查指出，Koha 是全美國排名第一的全功能的開放原始碼圖書館管理系統，擁有最大的圖書館市場，主要在公共圖書館這個區塊。Koha 在英國和台灣的市場尚在萌芽階段，未來在這兩個國家仍有發展空間(Bissels,

2008)。

2. Koha UNIMARC/CMARC/MARC21 測試平台

我們的測試平台是採用 Koha 當作圖書館管理系統，研究 UNIMARC、CMARC 和 MARC21 分欄特性。書目資料是由葡萄牙國家圖書館(Biblioteca Nacional de Portugal) 提供 24 筆符合 ISO2709 的 UNIMARC 書目記錄，以及國家高速網路與計算中心圖書館提供 CMARC 和 MARC21 各 100 筆書目記錄進行測試。

定義 Koha 的 MARC 架構

Koha 運作之前，系統提供工具用來定義 MARC 欄位和分欄架構。目前 Koha 官方僅提供 MARC21 和 UNIMARC 模板。就算是系統已經提供預設欄位架構，使用者還是可以依圖書館實際作業重新定義。台灣使用 Koha 的圖書館大都使用 Koha-Taiwan 團隊發展的免費/自由的 CMARC 模板（毛慶禎，2006 年）。圖 1 是一個 CMARC 重新定義的例子。左側顯示的是預設的 700 欄位和分欄；而在右邊，一個分欄 \$s 已經被重新定義並且顯示匯入的書目記錄。即使沒有被定義，分欄中的資料仍然會完整保留在資料庫中，不會因為 MARC 架構沒有被定義到而導致資料遺失。

MARC biblio : 11 (黃金時代) MARC biblio : 11 (黃金時代)

With Framework : Default

0 1 2 6 7 8

700 #1 - 人名 -- 主要著者
a 標目主體 司馬
b 副標目 光
4 著作方式 原著

702 #1 - 人名 -- 其他著者
a 標目主體 柏楊
4 著作方式 譯

With Framework : Default

0 1 2 6 7 8

700 #1 - 人名 -- 主要著者
s 朝代 宋
a 標目主體 司馬
b 副標目 光
4 著作方式 原著

702 #1 - 人名 -- 其他著者
a 標目主體 柏楊
4 著作方式 譯

圖 1：重新定義 CMARC 分欄架構

我們的測試平台是根據 2008 年的 UNIMARC 第 3 版，2001 年 CMARC 第四版更新版以及 MARC21 的 2008 年更新版。

匯入 MARC 書目記錄

我們在匯入 MARC 書目記錄時沒有遇到任何困難，三種機讀編目格式的 MARC 資料都可以正確匯入；但是 Koha 預設的編碼系統是 Unicode（也就是 UTF-8），並不是台灣最通行的 CCCII 或是 Big5。如果不經過特別程式處理而直接匯入，則會導致亂碼的情形發生，無法正確顯示中文書目資料（蔡育欽，2007 年）。圖 2 顯示匯入後的 UNIMARC 書目資料。點擊最右邊的書目號碼，系統會顯示 MARC 書目資料。

Manage Staged MARC Records › Batch 3

| | |
|--|-----------------------------|
| File name | nchc_0513_chi.iso |
| Comments | import marc from nchc(none) |
| Staged | 2009-03-17 15:21:53 |
| Status | staged |
| Matching rule applied | No matching rule in effect |
| Action if matching record found | create_new |
| Action if no match found | create_new |
| Item processing | ignore |

| | |
|--|------------------------------------|
| New matching rule | Do not look for matching records ▼ |
| Action if matching record found | Add incoming record ▼ |
| Action if no match found | Add incoming record ▼ |
| Item processing | Ignore items ▼ |

Apply different matching rule

Import into catalog

Page 1 [2](#) [3](#) [4](#)

| # | Citation | Status | Match? | Bib |
|----|---|----------|----------|--------------------|
| 1 | Statistical theory of heat : Brenig, Wilhelm. (3540510362) | imported | no_match | 1 |
| 2 | Computational methods for kinetic models of magnetically confined plasmas / (0387134018) | imported | no_match | 2 |
| 3 | Linear kinetic theory and particle transport in stochastic mixtures / Pomraning, G. C. | imported | no_match | 3 |
| 4 | Physics of space plasmas : Parks, George K. (0201508214) | imported | no_match | 4 |
| 5 | A computational approach to chemistry / Hirst, David M. (0632024313 0632027436) | imported | no_match | 5 |
| 6 | Topological methods in chemistry / Merrifield, Richard E. (0471838179) | imported | no_match | 6 |
| 7 | Algorithms for chemists / Zupan, Jure. (0471921734) | imported | no_match | 7 |
| 8 | Factor analysis in chemistry / Malinowski, Edmund R. (0471530093) | imported | no_match | 8 |
| 9 | Theoretical and computational models for organic chemistry / (0792313143) | imported | no_match | 9 |
| 10 | Mathematical frontiers in computational chemical physics / (0387967826) | imported | no_match | 10 |

圖 2：UNIMARC 書目記錄

MARC 書目全資料和簡易格式瀏覽

在瀏覽 MARC 書目資料之前，我們先定義好要顯示的格式、欄位和分欄書目資料。書目資料可以以國際書目著錄標準格式(ISBD)或是簡易格式顯示。這兩種格式都提供圖書館員和 OPAC 瀏覽界面。圖 3 是一個 UNIMARC 書目記錄顯示畫面。系統提供三種檢視書目資料的模式，分別為：一般檢視(Normal View)、MARC 檢視(MARC View)以及國際書目著錄標準格式檢視(ISBD View)。



圖 3：UNIMARC 書目資料 ISBD 檢視模式

語言編碼和傳輸

在處理 Koha 系統匯入資料時，重要的是具備有基本的資料編碼知識，因為不同的編碼方式會影響資料的處理及交換。ISO2709 定義欄位的長度，但是欄位的長度又根據不同字元編碼而有所差異。例如西文資料長度以字母為單位，每個字元為 1 byte，而傳統的中文資料如 CCCII 與 Big5 以字為單位。Big5 每個字為 2 bytes，CCCII 每個字為 3 bytes，因此需要額外特別程式處理。這種情況常見於亞洲國家如中國、日本和韓國。

我們的測試平台共測試了三種機讀編目格式。結果證明只要調整 MARC 欄位架構，任何 MARC 資料都可以正確匯入 Koha 系統。經過欄位調整後，可以完整顯示所匯入的資料。

3. 比較和評估

綜觀不同的機讀編目格式之間有許多微妙的差異，但是他們都維持著非常密切的關係，因為他們彼此都遵循英美編目規則。UNIMARC 的發展理念是從圖書館卡片目錄而來，MARC 欄位順序一如卡片目錄，從主要款目、敘述項、主題款目到其他附加款目（追尋項）等。

UNIMARC 的欄位層次安排較合乎功能邏輯。先是識別段/國際標準編號和代碼資料段（MARC21 則開始於代碼資料段和識別段），然後是依國際書目著錄標

準格式的敘述項、附註段、連接款目段（這裏可能是附註段，因為傳統的連接款目段是以附註表示）、主題分析段，接著是 700 著者段。即使是在同一個區塊 UNIMARC 和 MARC21 的欄位不盡相同，例如 6xx 主題分析段，UNIMARC 欄位 601 是團體名稱標題，而在 MARC21 是 610。然而，這些差異都是細微小事。就相容性而言，更重要的議題是有關分欄細膩深度以及連接款目段欄位資料連接的問題。因此，就執行系統軟體而言，相關的代碼資料段欄位以及連接款目段欄位就需要個別測試評估他們的功能特性。

將連接款目段欄位資料連接到其他欄位是一個複雜的程序。就資料著錄而言，代碼資料段欄位編碼極為錯綜複雜。這是因為資料代碼欄位需要仔細計算編碼長度才能正確著錄資料，否則整個資料著錄程序會是相當棘手的，因此我們需要一套量身定做的處理程序法則。資料代碼欄對於檢索多國語文/多國字體資料有非常重要的關鍵性。這些資料儲存在資料代碼欄是最適切的。就如同透過定義網頁語言以求提高搜索引擎檢索精確率和定義編目語言求提高檢索精確率是同樣重要的。此外，使用者通常相當關心資料的語言，甚至也對原著資料語言有興趣。這些有關資料的語言資訊都儲存在資料代碼欄。如果系統不能處理多國字體的問題，則需要經過適當的程式處理，以便系統將資料正確呈現得以被瀏覽檢索。

比較

A. 代碼資料段

我們測試了 UNIMARC 和 CMARC 代碼資料段欄位，也就是定長欄位 100、105、110、115、116、117、120、121、125、126、128、129、130、135、140 和 141；也測試了 MARC21 代碼資料段欄位 007 和 008。我們發現 Koha 系統只預設定義到 UNIMARC 和 CMARC 的欄位 100 和 105 的分欄 \$a 如表 1；但是 MARC21 欄位 007 和 008 所有的分欄卻都已經預設在系統如圖 4 所示，點擊圖的最右邊就會顯示全部的分欄。

這意味著如果使用 UNIMARC 或是 CMARC 系統館員必須重新定義 Koha 欄位架構；如果是使用 MARC21 則不需要。就技術支援層面來看，這是有利於 MARC21 的，因為圖書館可以省去額外的定義欄位架構的工作。

| tagfield | tagsubfield | librarian |
|----------|-------------|-------------------------|
| 100 | A | General Processing Data |
| 105 | A | Monograph Coded Data |

表 1：UNIMARC 預設分欄


```
<datafield tag="481" ind1=" " ind2="1">
  <subfield code="1">001295852</subfield>
  <subfield code="a">Loura, Luis Armando de</subfield>
  <subfield code="c">Lisboa. -[s.n.,</subfield>
  <subfield code="d">D.L. 1956]</subfield>
  <subfield code="t">Regulacao de avaria grossa pela arribada a Durban
    do N. M. "Benguela", em 12 de Agosto de 1952</subfield>
  <subfield code="5">PTBN: S.A. 5516//3 A.</subfield>
</datafield>
```

評析

正如我們在第 3 節討論的，雖然 UNIMARC 的欄位層次安排比 MARC21 更合乎功能邏輯，但是從圖書館管理系統的角度來看，我們測試的結果顯示使用 MARC21 的 Koha 圖書館處於比較優勢的立足點。這是因為最重要的資料代碼欄在 MARC21 欄位 007 和 008 已經預設定義好；而在 UNIMARC 和 CMARC 資料代碼欄散佈在多達 10 多個尚未預設的欄位架構之中。也許這就解釋了為什麼比起世界其他國家，Koha 系統在美國市場穩步成長的原因。

從圖書館的角度來看，如果圖書館有足夠的技術資源，可能不會選擇使用免費的圖書館自動化系統，但是如果圖書館沒有足夠的資訊人員，而且選擇使用 Koha 系統，可能會優先考慮 MARC21。雖然 Koha 系統沒有預設 MARC21 連接款目段欄位，我們之前提到，圖書館仍然可以根據實際作業，定義所需要的欄位架構。

4. 結論

Koha 是一個成熟且具有許多優點的圖書館自動化系統。誠然，Koha 提供 MARC21 和 UNIMARC 兩種預設模板。這意味著 Koha 的設計比較適用於 MARC21 或 UNIMARC；使用 CMARC、JAPAN/MARC 或是 Korean MARC 則需要特別的程式處理，書目資料才能正確呈現。對於資訊產業較不發達的國家來說，龐大的圖書館系統技術維護工作是很複雜的，許多發展中國家的圖書館並沒有足夠的能力做到這點。

美國是 MARC21 格式大本營，Koha 系統在美國市場大幅成長是可以理解的。英國現在很多圖書館已經轉向使用 MARC21，有些圖書館仍然遵循與 UNIMARC 格式細膩度相近的 UKMARC，這可能是 Koha 系統直到最近幾年才開始出現在英國市場的主要原因。台灣普遍使用的 CMARC 其格式細膩度與 UNIMARC 極盡相似。雖然 Koha-Taiwan 團隊持續義務提供 Koha 社群實際而且有用的服務與

資訊，唯社群規模仍然不大。

我們的研究結論是，Koha 系統使用 UNIMARC、MARC21 或是任何其他的機讀編目格式搭配商業圖書館服務公司的支援，如同 Koha 系統在美國市場的商業模式是我們建議的最佳方案。

參考文獻：

- Bissels, Gerhard. (2008). Implementation of an Open Source Library Management System Experiences with Koha 3.0 at the Royal London Homoeopathic Hospital, *Program: electronic library and information systems*, 42 (3):303 – 314.
- Breeding, Marshall. (2007). An Update on Open Source ILS, *Computers in Libraries*, 27(3): 27-29.
- Breeding, Marshall. (2008). The Viability of Open Source ILS, *ASIS&T Bulletin*, 35(2): 20-25.
- 毛慶禎 (2006)。Koha Taiwan。Google Groups。線上檢索日期：2009年3月22日。網址：<http://groups.google.com/group/kohataiwan>。
- 毛慶禎、許靜芬 (2006)。中國機讀編目格式(台灣)與書目資料庫。第72屆國際圖書館協會聯盟年會及世界圖書館資訊大會年會，2006年8月20日至24日，韓國首爾。
- 蔡育欽 (2007)。/trunk/cmarc/marc_to_utf8.pl - Koha Taiwan - Trac。
http://trac.koha-tw.org/koha/browser/trunk/cmarc/marc_to_utf8.pl。
/trunk/cmarc/new_split.pl - Koha Taiwan - Trac。
http://trac.koha-tw.org/koha/browser/trunk/cmarc/new_split.pl。

原文刊載於第75屆國際圖書館協會聯盟年會及世界圖書館資訊大會年會，2009年8月23日至27日，義大利米蘭。

誌謝

許靜芬 國家圖書館書目資訊中心編輯

大同大學 98 學年度基礎案