



聆聽「Metadata發展現況：資源描述與主題檢索」 演講有感

陳微麗

隨著電腦與網路的結合，使得網際網路的應用日漸普遍，資訊藉此快速地傳播。然而在充滿各式資源的網路世界中，唯有將豐富的網路資源予以組織整理，方能使之成為有用的資源。圖書資訊學界著名學者麥麟屏教授於5月20日應臺大圖書館、臺大圖書資訊學系及中華圖書資訊館際合作協會北部分區委員會之邀，假臺大圖書館總館地下一樓國際會議廳，以網路資源的組織整理之相關議題即「Metadata之發展現況」為主題，發表專題演講，茲將演講的精彩內容加以整理如下，以饗讀者：

網際網路的相關技術不斷地推陳出新，使得網際網路的環境持續在進步與發展，而由於網路環境具有下列特性，對圖書館在資料的組織與整理方面產生了若干影響：

1. 多元性：由於網路的快速傳遞，資料的數量愈來愈龐大，類型不再侷限於某一種類而趨向多元化，使得資料的組織整理無法以同一種類型的處理原則予以整理。
2. 流動性大且不穩定：分散式的網路架構使得資源的儲存分散在各處，由於資料並非圖書館本身的實體館藏，很難記錄其確切的館藏地。且可能隨時改變儲存的位置及更新版本，使資料來源難以掌握。而由於網路資源是無形的，以致很難定義資料的類型並予以歸類。
3. 複雜的版權問題。

此外，改為保存網路資源的版本，雖然可以解決紙本的缺點，但使用上，軟硬體成為另一個新問題。亦即必須藉由電腦及網路相關設備才能使用豐富的網路資源。

為使網路上的資源能藉由系統化的方式加以組織整理，以方便大眾利用，發展一套統一的標準描述格式以符合所有的網路資源，即成為當前各界所關注的焦點。順應這樣的新觀念之趨勢而發展出Metadata，期望能由作者本人或非專業人員自行描述網路資源的基本特性。而這一套描述格式的標準包括以下幾個特色：

1. 通透性 (Interoperability)，可溝通不同組織系統：
2. 彈性 (Flexibility)，以因應不同資料特色的需求：
3. 適用性 (Adaptability)，能適應不同的環境及需要：
4. 擴展性 (Extensibility)，可視資料特色擴展所需項目：
5. Scalability，使用者可決定記錄的複雜程度或簡單程度：
6. 易用、易懂 (Simple to use and to comprehend)，以便一般人能自行描述資源。

為因應這樣的趨勢，麥教授認為圖書館界應重新分配工作，專業的館員應跳脫傳統的角色，改為從事相關標準的建立，例如：主題表、主題分析等，並利用電腦的功能來執行



重覆性質的工作。以下分別就資源描述及主題檢索兩方面介紹Metadata的發展現況：

一、資源描述

為有效地組織整理網路資源，各界因不同的目的發展出不同領域的Metadata以因應電子資源的整理，以下介紹幾種目前常用的標準：

1. AACR2R/MARC21：圖書館長久以來所發展的一套組織整理資料的編目規則及機讀格式，為因應電子資源的需求，分別更新改版。AACR2R之1998年的版本，於電腦檔案的章節，充分地將電子資源的新媒體和類型納入改版而成，並提出一份詳細的指南（LC Draft Interim Guidelines for Cataloging Electronic Resources）；MARC21，整合USMARC和CANMARC而成，並提供856欄位以記錄網路資源的網址。而OCLC的編目計畫之一InterCat，即是以AACR2R和MARC21為基礎所進行電子資源編目。

2. 都柏林核心集（Dublin Core）：創始於1995年，由圖書館、電腦、網路方面的學者和專家在OCLC及NCSA所聯合贊助的研究會中共同討論而產生的。目的是要讓資源的創造者或出版者等非專業人員能自行建立資源的Metadata，以利資訊檢索。

(1)原則：藉由國際性一致的原則（International Consensus），以發展的一套描述網路資源之格式，因此簡單性（Simplicity）為其主要原則；而為使網路上各種不同的資源描述格式類型，能藉由一定的原則以達成精確檢索的目標，且期使同一領域中的語意互通（Semantic Interoperability），因此語意處理上採用相關標準來建立資料；此外並保持擴展性（Extensibility）原則，以符合不同需求、因應不同資源特性。

(2)項目：共制定15個欄位，包括：題名Title、著者Creator、主題Subject、相關敘述Description、出版者Publisher、其他參與者Contributor、出版日期Date、資源類型Type、資料格式Format、資源識別號Identifier、資源來源Source、語言Language、關連資料Relation、內容範圍Coverage、使用權限Rights。

(3)應用：

a. GEM（Gateway to Educational Material）：協助中小學教材進行的資源描述格式，將重要但未編目的教育資料予以組織整理，以提供教育者更快速及方便的檢索。

b. SignPost：有助於研究人員及教員使用。

c. CORC：由OCLC與全球圖書館合作的全球性計畫，合作建立高品質及為圖書館選擇以web為主的電子資源資料庫。

3. 其他類型的Metadata：

(1)TEI（Text encoding initiative）文獻編碼計畫：為國際性的計畫，以促進資源的交換為目的，發展一套指導方針，做為電子文獻交換及再利用的準備。規定每篇文獻之前，均應放置所有描述該文獻（包括來源、版本等）的標示文字（即文獻的Metadata）。包括：



- a. 說明文獻之書目特性及出處 (File description) :
- b. 編輯過程之決定的記錄 (Encoding description) :
- c. 描述語言、參與者、主題等非書目性資料 (Profile description) :
- d. 更新、修訂記錄 (Revision description) 。

(2)EAD (Encoded archival description) : 國際性的標準, 目的在於提供檔案、手稿資料一個標準的機讀描述格式, 並促進網路上手稿、檔案資料的取得。以SGML為基礎但亦可適應XML。受到MARC的影響, 提供階層式的、更詳細的描述, 並具數位項目連結的特性。始於1993年加州柏克萊大學 (University of California, Berkeley) 的計劃。包括

- a. 檔案描述或來源 (archival description) :
- b. 相關的出版資訊, 例如: 書名頁和其他序言部分 (frontmatter) :
- c. EAD的核心部分, 包括檔案本身的描述 (archdesc) 。

(3)GILS (Government information locator service) 政府資訊指引服務: 為方便一般大眾及政府職員取得政府機構相關的資訊, 由美國聯邦政府所建立以描述政府資料的標準, 受到MARC的影響並採Z39.50標準建立分散式架構, 使用SGML及MARC的格式來組織資訊。所包含的項目超過150個, 包括資訊的來源、管理、著作權及檢索限制, 項目及欄位的使用均採用MARC語法。

二、主題檢索

主題分析對資源的組織整理而言是相當重要的一環, 可使檢索者依不同的主題對已組織的資源加以檢索利用。目前美國圖書館界對Metadata的主題檢索成立了專門的小組委員會 (ALCTS/SAS/Subcommittee on Metadata and Subject Analysis), 研究探討在使用Metadata進行數位資源的分類及主題分析時所產生的主要問題, 並針對這些問題提供相關的討論群 (discussion forums) 及計畫。上述的主要問題包括:

1. 控制辭彙的優點:
2. 分類資料的使用:
3. 結構性的問題: 語意 (Semantics)、語法 (Syntax)、詞彙關係 (Term relationships) :
4. 系統設計的問題:
5. 應用上的問題: 修改現存的及重整或發展新系統 (Existing vs. modified or new schemes)、應用原則、維護。

(一) 小組委員會針對上述問題分別就詞彙及系統設計方面提出建議:

1. 詞彙方面:

- (1) 關鍵字和控制詞彙應合併使用: 讓使用者能以類似複雜的控制詞彙索引般做 free-text 全文索引的選擇。



- (2) 採用已存在的詞彙：例如LCSH或修飾語、未修飾語，以使語意一致，且可使一般性的字彙能包含所有的主題。
- (3) 採用簡單的語法：Metadata 記錄中，尤其是主題方面的欄位，應依使用者需要以決定其簡單及複雜程度。而對未受過主題編目訓練的人員而言，簡單的語法是最適合的方式。可行的方法包括：
 - a. 從語意上分離語法--從主題結構將詞彙的語意予以分離，例如應用LCSH的語法，斷開無主題的項目（地區、年代、形式複分）。
 - b. 設計一系列彈性的語法，以提供不同程度的使用者。
 - c. 前組合和後組合的應用。
- (4) 整合不同控制詞彙系統的主題詞：發展一套一般性的metathesaurus或鼓勵擴展醫學的metathesaurus，以包含各種主題領域的索引詞彙。

2.系統設計方面：

- (1) 使用自然語言的處理方式，依據全文的字彙以自動產生關鍵字索引。
- (2) 以資源本身為基礎，自動分類資料。
- (3) 從資料相似的项目，將主題和分類的資料自動濾出。
- (4) 可直接在線上檢索到控制字彙和分類表。
- (5) 可直接將使用者輸入的字彙與控制字彙和分類資料自動進行比對的作業。
- (6) 可藉由線上的工具或協助功能提供非編目員適當的主題詞彙和分類號。例如，提供線上權威控制或列出控制辭彙：設計藉由主題詞的下拉式選單直接給予主題詞及連結author-input 詞以控制辭彙的介面。

(二)小組委員提出目前採用的幾種主題檢索表並分別予以介紹：

- 1.LCSH：採用LCSH豐富的字彙，以簡單的語法，從各個不同層面來取代一個字串，使一般人均能自行描述網路資源的主題。另一方面基於語意及語法的考量，前組合在使用上，能以完整的主題字串做為一般Metadata的主題資料。而由於後組合具簡單的主題標題結構、簡單的索引展示、簡單的權威控制，並具容易使用及了解且可結合其他的控制辭彙的優點，因此對於系統未提供完整字串的瀏覽和搜尋的參考功能，及受過短期專業訓練的人員而言，可給予有關網路搜尋引擎和自然網路資源檢索的特質。
- 2.MeSH：Locator Plus採用MeSH，以為主題檢索之詞彙。
- 3.FAST（Faceted application of subject terminology）：結合CORC的研究計畫，採用LCSH的詞彙，為後組合面向的字彙，專為線上環境並為方便受過部分專門訓練及具有經驗的人而設計。Facets所包含的項目為：主題（人名）：地區（組織名）：形式（會議）：時間（uniform titles）。



- * 應用模式 LCSH Topical Headings (650)
 Geographic Headings (651)
- Faceting ↓
- FAST Topical Headings
 Geographic Headings
 Form Headings
 Period Headings

< 範例 > Church architecture \$z Germany \$z Rhineland \$v Exhibitions.
Architecture, Modern \$y 20th century \$z Germany
\$z Rhineland \$v Exhibitions.



- Topical: Church architecture;
 Architecture, Modern
- Geographic: Germany-Rhineland
- Period: 1900-1999
- Form: Exhibitions

網路的發展為圖書館帶來了各方面的衝擊及變革。傳統的組織及整理的方式已無法有效地處理這些電子式的網路資源，且由於網路資源的氾濫，實在無法一一地加以組織整理，因此如何選擇便成為重要的課題，而資料的數位化也改變了傳統的保存方式，面對這一連串的問題，圖書館必須開始重新思考整個定位與作業方式。

在資源的組織與整理方面，面對豐富的網路資源，如何有效地加以組織使之成為有用的資訊，已成為各界所關注的議題，也紛紛提出不同的計畫、作法，而對視組織資料為己任的圖書館而言，整合各界的作法並發展一套統一的標準描述格式，以符合所有網路資源，即成為當前最重要的目標。同時，用全新的觀念重新看待圖書館的作業方式，並調整各項工作的分配及角色的扮演更是我們所應努力的方向。

參考書目：

1. 麥麟屏，「Metadata發展現況：資源描述與主題檢索」演講大綱。
2. 陳雪華，「網路資源組織與Metadata之發展」，圖書館學刊第十二期（民國86年12月），頁19-37。
3. 陳昭珍，「電子資訊的組織模式」，圖書館學刊第十二期（民國86年12月），頁161-182。
4. 吳政敏，「都柏林核心集的發展現況與其在圖書館的應用」，網際網路與圖書館發展研討會論文集（民國88年12月），頁113-135。