



Metadata 概說

楊雅勳

資訊的載體、處理方法與技術，隨著科技進步、環境變遷產生極大的變化，愈來愈多人習慣透過網路查尋資料，然而網路資源的種類及數量眾多，為方便使用者檢索與利用各式資源，必須有效的組織整理網路資源等電子資訊，因此圖書館員應該注意及關心目前電子資訊組織模式的發展。

中國圖書館學會為因應此一趨勢，在今年所舉辦的「主題分析與機讀格式專題研習班」中邀請吳政叡老師介紹目前的熱門議題「都柏林核心集與元資料」。短短兩個小時的專題演講梗概介紹 metadata (元資料) 的定義、興起背景及都柏林核心集的發展，為了能對 metadata 有更完整、客觀的概念，所以筆者收集一些相關資料，試圖對 metadata 的種類與用途及未來趨勢做進一步的了解。

Metadata 常見的英文定義是 "data about data" 或者 "data describes other data"，意即描述資料屬性的資訊，國內常譯為詮釋資料、資源描述格式、超資料、元資料。metadata 對網路資源的內容、特質與儲存位置的描述，有助於使用者檢索、評估及選擇資訊。

為了各種不同的目的與用途產生的 metadata 相當繁多，Lorcan Dempsey 與 Rachel Heery 依 metadata 的架構、完整性及專業性，將之分為下列三大類：(註 1-3)

第一類：一般性的網路查尋工具，包括相當多無結構化的資料，且資料少有明確的語意，亦不支援欄位檢索，當使用者利用此類工具檢索時，往往是回收率大，精確率低，難以判斷大量的查尋結果究竟有多少和需求的主题確切相關。即使如此，網路上的檢索引擎、主題目錄仍被廣泛使用，因其目的在做全域索引服務，在檢索工具、服務上也不斷推陳出新(例如：使用內容式資訊過濾技術，提供個人化資訊服務等)，以吸引使用者利用。Yahoo、Lycos、Alta vista 即屬此類。

第二類：以蒐尋為目的的 metadata，包括足夠的描述，讓使用者在檢索後能判斷是否需要獲得全文。屬於結構化的資料，支援欄位檢索，例如 Dublin Core。Dublin Core 是一套簡易的資料描述格式(較簡單有彈性)，1995 年 3 月 OCLC 和 NCSA 所聯合贊助的研討會中，由圖書館、電腦、網路方面的學者和專家共同討論下的產物，目的是讓資訊提供者能對提供的資料加以描述，其簡易有彈性的資料描述格式易為各種專業人員所了解，滿足大多數非圖書館專業人員的需求。Dublin Core 有 15 個基本項目：主題(subject)、題名(title)、著者(creator)、簡述(description)、出版者(publisher)、其他參與者(contributors)、出版日期(date)、資源類型(type)、資料格式(format)、資源識別代號(identifier)、關連(relation)、來源(source)、語言(language)、涵蓋時空(coverage)、版權規範(rights)。其設計原理是使 metadata 的資料項，同時擁有意義明確、彈性和最小規模三種特色。所秉持的原則是：內在本質原則(intrinsicity)、易擴展原則(extensibility)、語法獨立原則(syntax-independence)、無必須項原則(optionality)、可重覆原則(repeatability)、可修飾原則(modifiability)。但 Dublin Core 目前尚在發展階段，因此有些議題未有定論，如修飾詞等(註 4，註 5)。除 Dublin Core 之外，IAFA templates、RFC 1807、SOIF、LDIF 等也屬此類 metadata。

第三類：以詳細記錄資源為目的的 metadata，資訊描述的完整性最高、最結構化，這類

metadata大多是為特定領域或特殊資料而訂定，且能詳細表達物件間的複雜關係，典型的例子即是 TEI Header、MARC、EAD、CIMI、ICPSR。

未來網路資源的組織整理及metadata的發展，尚未有一個十分明確的方向，但多數圖書資訊界與其他專業領域的人士對此相關問題做了許多探討，對未來的發展趨勢，做了下列預測：(註3，註5)

1. metadata 的格式：目前任何一個 metadata，均無法滿足所有使用者的需求。因此 metadata 將多元並存，且個別資料類型有其支配性的 metadata，例如美國聯邦地理資料委員會 (FGDC) 的地理元資料 (Digital Geospatial metadata)。此外，metadata 應具有包容性、易擴展性，以適應網路作業環境，任何一個 metadata 若具有排他性將無法久存。
2. metadata 的建立：各類 metadata 的目的不同，建立的方式也隨之不同。前述之第一、二類的 metadata 為彌補其資料不精確或改善速度慢的缺失，因此都在尋求作者或出版者提供資料的途徑，例如著錄項目的精簡和簡單好用。第三類 metadata 目的在做詳盡精細的描述，所以需要專家描述。
3. 通訊協定：在全球單一資訊網的作業環境下，沒有任何一台主機可以儲存和負荷所有的檢索需求，所以須採分散式的處理架構，並有相關的通訊協定配合，如 TCP/IP、Z39.50。

未來 metadata 的發展或許是多元並存，或許是每種 metadata 有其支配性的 metadata，不夠完備、不能適應環境的 metadata 便會被淘汰。圖書館界目前使用的 MARC，在整理資源上、應用技術上都很成熟，並且能將不同的資訊資源整合在一個目錄中，雖然使用 MARC 有許多優點，但是其他相關領域的專家仍覺得 MARC 不具彈性、不易使用等。未來我們希望能結合傳統的 MARC 與新興的網路組織模式，即結合編目與標誌，一方面解決 MARC 在應付電子資源上的困境，另一方面提供一套簡單的資源描述格式，讓眾多非圖書館界的專業人士使用，以最少的成本來解決電子資源快速增加的問題。另藉由 metadata 與 MARC 的對照轉換，將 metadata 資料轉入圖書館系統，再由館員豐富內容並做控制，以便使用者利用。

未來或許會發展出更正確、資料量更大、更完整、可隨時更新及處理多種語文能力的 metadata，我們應隨時注意國內外重要 metadata 格式的發展，並做出適當的回應。

參考文獻

- 註 1. Lorcan Dempsey and Rachel Heery, " Metadata : an Overview of Current Resource Description Practice" , <<http://www.ukoln.ac.uk/metadata/DESIRE/overview/>> 。
- 註 2. 陳雪華，「網路資源組織與 metadata 之發展」，圖書館學刊 12 期(民國 86 年 12 月)，頁 19-37。
- 註 3. 陳昭珍，「電子資訊的組織模式」，圖書館學刊 12 期(民國 86 年 12 月)，頁 161-182。
- 註 4. 吳政叡，「元資料實驗系統和都柏林核心集的發展趨勢」，國立中央圖書館臺灣分館館刊 4 卷 2 期(民國 86 年 12 月)，頁 11-24。
- 註 5. 吳政叡，「從元資料看未來資料著錄的發展趨勢」，資訊傳播與圖書館學 4 卷 2 期(民國 86 年 12 月)，頁 42-51。