
Metadata 初探

王麗蕉

一、前言

隨著電腦科技與網路通信的進步，人類社會邁進了嶄新的資訊網路時代，世界各先進國家紛紛投入資訊基礎建設的行列，積極建設資訊高速公路，以提振在資訊網路時代國家的競爭優勢。為充實資訊網路內涵，電子圖書館相關計畫普遍受到各界重視，而在建構電子圖書館一個必要過程是將現有的館藏、出版品經過電子化處理，並提供資訊的檢索、查詢、瀏覽、以及傳遞。在電子化轉換的過程，則需要一適合在網路環境中有效管理電子資源、並能適切呈現不同館藏特色的資源描述格式，由於傳統機讀編目格式無法適用於大量網路電子資源的組織方式，為配合考量使用者的新資訊需求等因素，metadata，這個新的電子資源描述格式(或稱電子資源組織模式)，已成為發展電子圖書館服務計畫時，有關資源管理中首要考量的問題。本文接著將就 metadata 電子資源描述格式的定義、發展背景、與種類，以及 metadata 發展所面臨的問題作一簡略的介紹。

二、metadata 的定義

何謂 metadata? 在英文文獻中，最常見的解釋為「data about data」，直譯成中文則為「有關資料的資料」或「描述資料的資料」。國內學者針對 metadata 一詞，則有吳政叡教授所稱之「元資料」，陳雪華教授的「詮釋資料」，陳昭珍教授稱之的「超資料」。有關 metadata 進一步的解釋則有：

Priscilla Caplan¹ :

*Metadata 即是描述資料的資料，編目記錄就是 metadata ; TEI (Text Encoding Initiative) header，或任何其他描述形式都是 metadata。
(Metadata really is nothing more than data about data, a catalog record is metadata; so is a TEI header, or any other form of description)*

國際圖書館協會聯盟 (International Federation of Library Associations and Institutions, 簡稱 IFLA) 的定義ⁱⁱ：

metadata 即描述資料的資料，metadata 一詞意謂可用來協助對網路電子資源的辨識、描述、與指示其位置的任何資料。(Metadata is data about data. The term refers to any data used to aid the identification, description and location of networked electronic resources)

從 Caplan 對 metadata 解釋中，編目記錄就是 metadata，則圖書分類、編目等傳統圖書館資料組織的方法，可描述圖書資料的內容、特色或資料實體 (例如作者、題名、主題、館藏位置等)，進而提供協助資料檢索之功能的分類編目，其實已是 metadata 的明顯應用了。此外，從 IFLA 的定義來看，隨著資訊網路的發展，大量電子資源的生產以及使用者資料使用環境的改變，期盼透過網路可一次查檢並取用所需資源，因而產生了新的資訊需求，所以新發展的資源描述格式(metadata)，用以協助辨識、描述與網路電子資源以及指示資源所在的位置，metadata，成為目前各界所熱衷發展的新資訊組織方式。

歸納以上的解釋得知，所謂 metadata，廣義而言，是指對任何資料(泛指圖書資料、非書資料、網路資源、任何形式的電子資料)所描述的任何資料(書名、主題、外在表徵、位置)。狹義的解釋為：用來定義、辨識電子資源、以及協助資源取用的描述方式。為了與傳統分類編目有所區分，並應用於現今網路環境中的資源組織方式發展趨勢，本文就 metadata 給一較狹義的名稱，稱之為「電子資源描述格式」。

Metadata(電子資源描述格式)包含的資訊很多，例如題名、作者、出版者、主題、歷史沿革、檢索控制、使用權限、與其他作品的關係、適用對象、以及相關內容範圍等都可涵蓋其中。因此，metadata 除了兼具傳統書目工具匯集與辨識功能外，還具備下列特點：1.定位(location):提供所需資源的儲存位置;2.探索(discovery):提供使用者檢索到所需的資源;3.文件記錄(documentation):提供詳細描述與記錄文件的性質與內涵;4.評估(evaluation):協助使用者判斷資源是否具有價值;5.選擇(selection):幫助使用者決定是否取用此資源ⁱⁱⁱ。

三、電子資源的知能控制與 metadata 發展背景

由於資訊時代網路的發展，促使圖書館將館藏資源電子化，以此作為建立電子圖書館系統的基礎，成為目前圖書館提供資源服務的新發展趨勢之一。建構電子化館藏的進行方式，主要有三種：(1) 將原有的館藏、出版品經過電子化處理，轉換成電子化資源；(2) 購買電子化的出版品，由於電子出版技術的進步，產生了各式的電子出版品，例如電子期刊、電子圖書、以及 CD-ROM 等電子化多媒體；(3) 蒐集、整理網路上大量產生各種類型、內容豐富的網路資源。

然而，不論以何種方式進行，電子化資源是發展電子圖書館的必要條件，也是電子圖書館主要的特色之一。如同傳統圖書館的書目控制工作，在電子圖書館中，如何建立、描述電子化資源，以匯集與辨識資源的特質與內涵，提供有價值利用的資訊給資訊網路中的使用者，為此，必須發展適合電子資源描述、組織的方式，才有效管理大量發展的電子資源，作為電子資訊查尋、檢索、交換與管理等知能控制的基礎。

在傳統印刷式出版品的環境，國際書目控制可藉由出版紙本式的國家聯合目錄達成，接著由於自動化系統的發展，可以經由線上聯合目錄，提供全球性檢索，達到國際書目控制的目的，促使國際間的圖書資源得以共享。然而在網路電子資源大量出現，全新的電子出版模式，造成各式電子資源成長的速度、出版的形式種類超乎人的想像之外；傳統圖書館的書目控制工具，例如機讀編目格式、分類方法等傳統資料組織方式，無法適用於網路出版環境的電子資源組織與整理，以符合使用者的資訊需求，因而必須發展新的 metadata 資源描述格式，來組織、整理電子資源，以提供電子資源的檢索與利用，達到電子資源知能控制的目的。

總括而言，發展 metadata 有其外在的環境因素，茲分述如下：

(一) 電子資源的多變與多元化的特色

電子資源很容易複製與更改，造成版本控制不易；電子資源的存放位置也很容易更動，有些只是短暫存在，有些也因資源更新連帶更動了存放的位置；電子資源的種類多元化，不像印刷式資料有固定的格式存在，一份電子資源可能包括數個檔案，分放在不同的系統或機器上，這些電子資源的特性，造成利用傳統資料分類編目方式整理電子資源的困難度。

(二) 傳統資料組織模式不適用於特殊藏品的電子化之描述

電子圖書館的第一步驟是將特定藏品經過電子化處理，傳統的編目規則或機讀格式無法完整詳細描述特定藏品的屬性、特徵以及物件間的相互關係，故必須發展適合特殊藏品電子化資源描述 metadata 標準，符合文獻特性與使用者之資訊需求行為。而圖書館界使用而久的機讀格式(MARC)不適用於作為特殊藏品電子化描述之依據的因素為：(1)MARC 的記錄有長度之限制，最長為 100,000 個位元；(2)MARC 無法完整的描述階層式資源，例如檔案資源；(3)MARC 的市場環境較小^{iv}。因此，發展新的資源描述格式，除了應符合文獻特性，滿足使用者之需求外，也應考量到行政與管理的問題。

(三) 在提供資源服務時，產生新的競爭者

由於網際網路普及應用，網路上產生了大量電子化資源，許多機構與個人設計了各式各樣的網路資源組織與管理之查尋系統，例如資源指南目錄(Yahoo, 蕃薯藤)、搜尋工具(Veronica, Lycos, Alta Vista)協助網路資源之查檢利用，此類資源查尋工具，目前大多以自動搜尋、自動索引方式，並透過相近元運算、Fuzzy search 以及相關性排序等檢索功能，提供快速、簡單的查尋功能；但此類搜尋工具一般採用自動搜尋、索引方式產生，對於收錄的資源品質良莠不齊，且未對資源屬性描述加以控制，容易產生檢索結果「誤引」的情況；而圖書館界一向以資源組織與管理專家自居，也瞭解到在電子圖書館時代，對網路資源組織整理的重要性，除了調整傳統機讀格式以配合電子資源的特性外，更積極投入開發新的電子資源描述標準的工作。

(四) 網路與電子圖書館的發展，產生新的資訊需求

在 IFLA 於一九九五年進行有關書目基本功能的分析計畫中，發現使用者對書目的需求有^v：

1. 「找尋」的需求：即能從檔案或資料庫中找到與所用之檢索條件相符的資料；
2. 「辨識」的需求：對於具有相同特性之資料，可辨識其間之差異；
3. 「選擇」的需求：即可根據內容、外形等描述條件，選擇適合其所需的資料；
4. 「取用」的需求：指可獲得檢索到的資料。

(五) 分散式資料庫互通之使用需求^{vi}：

由於網際網路的發展，使用者期盼能整合多個分散於不同系統平

台、不同資料庫檢索系統，藉由互通性的檢索介面能一次檢索並獲取到所需的資源，即所謂分散式資料庫互通之需求；資料庫互通的先決條件為解決「語意異質(semantic heterogeneity)」問題，所以必須發展共通的資源描述格式，來定義屬性命名、編碼格式、結構、資訊描述項目、描述依據以及主題分析、權威控制、字碼等。

此外，除了外在客觀環境的衝擊之外，metadata 在近幾年來蓬勃發展的原因，亦為 metadata 本身在網路上全球資訊檢索的重要性，其理由為：(1)metadata 能以較低頻寬網路來管理大量的電子資源；(2)藉由 metadata 元素與結構化的設計可深度分析資源的特質與內涵，有效協助複雜的電子資源之探索與，使用；(3)metadata 可整合異質的資訊資源，提供分散式資料庫的資源分享；(4)metadata 亦可處理有關檢索限制與限制使用的資訊安全管理，可以兼顧確保個人隱私，以及提供資源合理利用的雙重功能^{vii}。

外在客觀環境的衝擊，以及 metadata 內在條件的優勢，是以各個不同領域發展了各式 metadata 格式，例如文獻編碼計畫(Text Encoding Initiative, TEI)的 TEI Header、政府資訊系統(Government Information Locator Services, GILS)的 GILS Core、地理資訊的 CSDGM(Content Standards for Digital Geospatial)、博物館資訊的 CIMI(Consortium for the Interchange of Museum Information)、以及檔案資源的 EAD(Encoded Archival Description)等 metadata 標準，圖書館界亦積極發展一有效整理大量網路資源的都柏林核心(Dublin Core)。

四、metadata 的種類

目前 metadata 的種類因其發展的背景、目的以及描述處理的資源對象不同，而產生了各式各樣的電子資源描述格式，有些正積極朝向國際標準進行，例如都柏林核心(Dublin Core)，有些則已發展成國際標準，例如 EAD, TEI header 等；metadata 的發展不僅僅在圖書館界熱烈展開，在政府資訊、地理資訊、檔案界與博物館等不同領域，亦已發展或正在進行適合其領域資源描述所使用的 metadata 標準化的工作。Metadata 的種類繁多，根據 Locan Dempsey 與 Rachel Heery 於一九九七年三月發展的研究報告「Specification for Resource description Methods. Part I. A Review of Metadata: A Survey of Current Resource Description Formats」依 metadata 的結構性、完整性以及專業性分為三種：

第一類：直接採用原始資料（即未結構化的資料）的資源描述格式，一般性

的網路資源查尋工具、搜尋引擎皆屬此類，例如 Lycos, Yahoo, Alta Vista 等；

第二類：使用結構化資料，可支持欄位查詢，此類 metadata 描述資料結構簡單，可由非專業人士或資源創造者自行描述、著錄足以表示其文件資源屬性內容，主要目的在於提供對大量網路資源搜尋功能，而傾向對個別物件的描述，而不做多元化物件關係之描述，目前圖書館界積極發展的都柏林核心則屬此類；

第三類：使用完整的資源描述架構，用來記錄文件或描述一組物件、以及其彼此間相互關係，除了可提供定位與搜尋功能外，此類 metadata 完整詳細描述物件間的複雜關係，可表現特殊領域或特定文獻資源的特質與內涵，所以此類 metadata 需要具專業知識背景的人士才能建立與維護，方能滿足其專業需求，例如檔案採用的編碼標準 EAD、描述地理空間資源的 CSDGM、與描述博物館資源的 CIMI。

根據 Locan Dempsey 與 Rachel Heery 依 metadata 的結構化等特性所區分的三類之中，第一類採用未結構化資料的資源描述格式。此類資源描述方式通常採用自動搜尋資源、自動製作索引，雖可作為大量網路資源的快速查尋工具，但其未採用結構化資料作為描述依據，而資料項的格式並無一定之標準，除了會造成查詢上的誤引或盲點之外，也無法進行資源交流，故本文就資源有效管理之立場對此類 metadata 將不多作討論。以下將試就第二類簡單結構資源描述格式的代表 Dublin Core (都柏林核心)與第三類完整詳細描述格式的標準 EAD (檔案編碼標準)，簡述其發展背景、包含的描述元素項目、以及其具備的特質。

(一) 都柏林核心(Dublin Core)：

都柏林核心之發展，是為了便利於網路電子資源的探索，在一九九五年三月由國際圖書館電腦中心(Online Computer Library Center, 簡稱 OCLC)與國家高速電腦應用中心(National Center for Supercomputing Applications, 簡稱 NCSA)所聯合贊助，結合來自圖書館、電腦和網路界的專家學者共同研擬提出者，主要目的在於希望建立一套描述網路上電子資源的簡單格式，以增進網路資源檢索與利用。

經過多次研討會之討論，目前都柏林核心包括項資料項有十五項，分述如下^{viii}：

1. 題名(title):是由創造者或出版者給予資源的名稱；

2. 創造者(creator):資源知能內容創造的個人或組織機構；
3. 主題(subject):資源的主題，即敘述資源主題或內容的關鍵字或片語，包括控制詞彙或分類架構；
4. 簡述(descriptions):資源內容的文字性描述，包括文件的摘要或是影像資源的內概述；
5. 出版者(publisher):負責資源發表的組織，例如出版社、大學部門或團體組織；
6. 參與者(contributor):對資源的創造有相關貢獻的個人或組織，例如編者、譯者或插畫者；
7. 日期(date):資源發表呈現的日期；
8. 類型(type):資源的種類，例如首頁、小說、詩、技術報告、字典等，預計將有一資源資型的列舉清單可供選用；
9. 格式(format):資源的檔案格式，例如 text/html、ASCII、或是 JPEG 影像檔等；
10. 識別代碼(identifier):用來標示資源唯一的字串或數字，例如網路資源的 URL 或 URN，以及 ISBN 或其他正式名稱；
11. 出處來源(source):作品是由何處衍生而來的；
12. 語文(language):資源內容所採用的語文；
13. 關連(Relation):與其他資源的關連，例如所屬的系列或其他關係；
14. 涵蓋的範圍(Coverage):資源的時間與空間特性；
15. 版權(Rights):資源版權聲明以及版權管理使用之規範。

從都柏林的十五項資料項，可以發現其資源描述格式非常簡單易懂，此為是都柏林核心最初發展的概念--簡單化的設計原則，目的在於讓資源的創造者或出版者可以自行建立符合其資源內容的 metadata，而不需要具備深厚學科知識的專家來建立，而描述的資料可直接載入資源檢索指南以供瀏覽檢索，且結構化的資料，又可增進網路資源檢索的精確性與資源分享的目的。

(二) 檔案描述標準(Encoded Archival Description, 簡稱 EAD)：

EAD 的發展早自一九九三年起，由加州大學柏克萊分校(UC Berkeley)開始企圖發展一個適合檔案資源描述編碼的標準，接著由美國國會圖書館與美國檔案管理者協會(The Society of American Archivists)負責 EAD 的制定與維護工作，一九九六年發表的 EAD beta

版，經由幾個大學與相關機構的測試，發現 EAD 的檔案資源描述標準可呈現檔案與手稿原本的特性，而在一九九八年發表其 EAD 1.0 版。

EAD version 1.0 包含四件正式技術文件：(1)文件類型定義 (Document Type Definition, DTD)本身;(2)Tag Library;(3)EAD 使用指南;(4)EAD 工作小組的實際範例。EAD 符合檔案、手稿特藏的特質，提供結構化、多層式、完整詳細的描述規範。EAD 的文件類型定義 (DTD)概述如下^{ix}：

<ead>

<eadheader> 用來辨識查檢工具，包括

<eadid>唯一性的辨識號碼或代碼

<filedes> 查檢工具的書目性資訊

<profiledis>記錄查檢工具的使用的語文以及文件編碼創造者的資訊

<revisiondes>修訂的描述；

<frontmatter>提供題名頁與其他正式出版有關的資料描述，作為檢索用；

<archdesc>提供檔案單元主要結構項目之描述，包含下列幾項次元素

<did> 提供辨識的描述，包含描述檔案單元時所需的資訊，例如單元名稱、單元的日期、單元的範圍長度、產生者、檔案館、單元存放位置與一般附註等；

<add>附加的描述資料，包含書目、檔案計畫、索引、相關資料等次元素；

<admininfo>行政資訊，包含查檢限制、採訪資訊、其他可使用形式、典藏歷史、處理資訊、使用限制、以及附註等次元素；

<arrangement>編排情形

<bioghist>傳記與歷史資訊

<note>一般附註

<odd>其他描述資料

<organization>組織

<scopecontent>範圍與內容

<dsc>附屬成分之描述，底下又可包含多個成分描述,例如

<c01>、<c02>，而每個成分描述下又可包含辨識資料、行政資訊、控制查檢、範圍與內容等次元素。

由 EAD 的文件類型定義包含的元素(資料項)及其架構，可瞭解 EAD 檔案資源描述標準，適用於描述文件與非文件的資訊組織方式，也適用於多元化的網路環境，EAD 的定義包含了結構化的資料層級，可完整的描述檔案資源的層級關係，並能在查檢工具中反映描述資料本身的內容，可促進檔案電子資源的檢索效益。EAD 的設計是以 SGML DTD 為依據，提供檔案查檢工具詳盡與相關的多層級之描述標準，符合檔案特藏本身具備的層級關係，可表示層級間的互相關係並加以連結，亦可將某一特定層級的資訊結構移轉，並可支援特殊項目的索引與檢索。EAD 的描述格式除了適合檔案、手稿特藏多層級的著錄原則外，依據 SGML 發展的結構化資源描述標準，更能達成電子資源交流、檢索、取用的智能控制目標。

五、Metadata 發展面臨的問題

由於資訊網路的發展，引起各領域積極發展適合其領域資源描述的 metadata 標準，除了上述描述檔案完整詳細記錄的 EAD，適合大量網路電子資源組織的 Dublin Core 外，適合電子文件編碼的 TEI(Text Encoding Initiative)、專為博物館資訊交換的 CIMI(Consortium for the Interchange of Museum)、政府資訊的 GILS(Government Information Locator Service)與地理資訊的 CSDGM(Content Standards for Digital Geospatial Metadata)等皆已發展為該領域的國際標準格式。故在設計 metadata 格式標準時，如何制定描述資源的元素，以及如何規範元素中的屬性，是要採可快速管理大量資源之簡單結構化的資料項，還是要符合文獻特性制定完整詳細資源描述格式，將是在發展一個 metadata 標準時，首先面臨的第一個問題。

在電子圖書館時代，必須要整合各類資源提供使用者便於利用，各種 metadata 間的對照與轉換間的問題，包括前面所述簡單結構與完整結構轉換時，可能遺漏一些 metadata 元素(資料項)，以及要以何種 metadata 標準架構作為轉換時的依據等問題，將是未來整合各式 metadata 要面臨的另一個難題。

在圖書館界應用已久的機讀編目格式(MARC)，是一種發展成熟的 metadata 標準，然而 MARC 在組織整理電子資源上有其系統開發與太過繁複

等困難，但是電子圖書館是傳統圖書館服務的延伸，而不是一替代品，未來是否應整合兩種資源；或是以 MARC 作為轉換各種 metadata 的基礎，也是值得我們去研究的議題；目前已有一些有關 Dublin Core 整合到 UNIMARC, USMARC 等研究，但尚未有一定論。

六、結論

國內目前有許多關於電子資源描述 metadata 標準議題探討，然而實際去研擬制定一格式標準的，僅有「台大電子圖書館與博物館計畫」中針對古文書擬定的 metadata 格式^x，到目前為止雖已發展出一古文書 metadata 格式，但仍為暫定的架構初稿，尚不能稱之為一標準。所以關於國內 metadata 電子資源格式的訂定，有以下幾點建議：

- (一) 成立一 metadata 工作小組：由圖書館專業性學會組成一 metadata 工作小組，負責相關研究與建立 metadata 標準的行政與統籌工作。
- (二) 研究國外已設計完成或開發中的不同領域之 metadata，在中文環境的可行性，中文資源描述的適用性。
- (三) 設計符合國內各種藏品(例如圖書、期刊、檔案、博物、地理性空間資料等)的 metadata，且必須考量全球資訊互通的議題，評估和選用國際標準的資訊組織方式，例如國際標準標誌語言(SGML)等，作為 metadata 格式設計的原則，達成資源分享的目的。
- (四) 評估 metadata 與 MARC(機讀編目格式)間的整合問題：基於圖書館長久以來使用機讀編目格式的經驗與結果，存在已久的機讀記錄必須與 metadata 資源進行對映或轉換等整合工作。
- (五) 配合開發網路檢索協定標準(Z39.50)介面：在設計以描述資源為基礎的 metadata 格式的時，亦須審慎考量未來資訊檢索的方式，Z39.50 即為一可提供整合性的電子資源檢索介面與通道的國際標準，可基於電子資源描述標準的基礎，提供互通性的分散式資料庫檢索，達成有效資源交流與分享的目的。

註釋

- i Wool, Gregory "A Meditation on Metadata" The Serials Librarian. v.33:1/2(1998):168.
- ii Wool, Gregory "A Meditation on Metadata" The Serials Librarian. v.33:1/2(1998):168.
- iii 陳雪華，「網路資源與 Metadata 之發展」 圖書館學刊。第 12 期(民國 86 年 12 月)頁 19-37。
- iv Pitti, Daniel V. "Encoded Archival Description: the Development of an Encoding Standard for Archival Finding Aids" American Archivist 60(Summer 1997):275-276.
- v 陳昭珍，「從使用者需求與文獻特性看圖書館界資訊組織模式發展趨勢」 21 世紀資訊科學與技術的展望學術研討會論文集。(台北市:世新大學圖書資訊學系,民 87 年)頁 245。
- vi 陳昭珍，「電子資訊的組織模式」 圖書館學刊。第 12 期(民國 86 年 12 月)頁 164-165。
- vii Xu, Amanda "Metadata Conversion and the Library OPAC" The Serials Librarian. v.33:1/2(1998): 180-181.
- viii Weibel, Stuart "The Dublin Core: A Simple Content Description Model for Electronic Resource" Bulletin of the American Society for Information Science. v.24 n.1(OCT./NOV. 1997):11.
- ix Ruth, Janice E. "Encoded Archival Description: A Structural Overview" American Archivist v.60 (Summer 1997): 317-327.
- x 陳雪華，張玉欣，「台大電子圖書館與博物館數位化資源之組織」 圖書館學與資訊科學。23 卷 2 期(民國 86 年 12 月)頁 62-64。
-