

## 館長的話

### 試論數位圖書館中 metadata 的意義及其應用

館長 呂宗麟

#### 一、前言

近年來由於網路資訊科技快速進步，促使傳統圖書館的經營模式(傳統圖書館蒐藏的大都是紙本式的文獻資源)面對著嚴峻的挑戰，因此，數位圖書館的建立與發展，儼然成為當代圖書館界努力邁向的目標之一，然而，若將數位圖書館定義於「數位化的館藏」，乃是忽視了圖書館人員在技術服務與讀者服務上的努力<sup>1</sup>。

筆者相當同意美國學者 Cleveland 在其〈數位圖書館：定義、課題與挑戰〉(Digital Libraries：Definitions、Issues and Challenges)一文中對數位圖書館的界定，其認為：「數位圖書館絕非是一套單一完整的數位系統，而是一群異質系統或資源透過網路連接，並整合在單一的界面之下，這些異質系統或資源可能包含指向紙本或數位媒體的書目資料庫、索引與資料找尋工具、目錄、相片、電子期刊、數據資料庫……等；儘管這些資源或許存在於不同的系統或資料庫中，但對讀者而言，它們卻有如是存在於單一系統之內」<sup>2</sup>，誠如國內學者柯皓仁所說，Cleveland 這段話點出了數位圖書館的一個重要目標，即是如何將散佈在各地，獨立運作與維護的數位圖書館系統和資源整合在一起，讓讀者毋需了解在要取得的資源實際上是分散各地，還是整合在一個系統之上，換言之，讀者需透過單一的系統或界面，即可取得需要的所有資訊，要實踐前述目標，必須達成數位圖書館系統間的互通性<sup>3</sup>。

Metadata 概念出現的原因，正是為達成上述目標，以及用來協助對電子資源的辨識、描述以及指示其位置的任何資源，亦如傳統圖書館經常運用的機讀編目格式(Machine-readable catalog, MARC)或許就是一種最普遍且為人所知的 metadata，因為它可描述圖書資料的內容，特色或資料實體(如

---

<sup>1</sup> 柯皓仁等〈詮釋資料與數位圖書館系統互通性之探討〉，國立台灣大學《大學圖書館》第5卷第1期，2001.3，頁49。

<sup>2</sup> 同前註文，頁51。

<sup>3</sup> 同前註。

作者、題名、主題、館藏位置等)<sup>4</sup>大體而言，MARC 是二十世紀六十年代為印刷型(紙本)資源的描述而開發，它實際上是紙本資源的描述、存貯、管理檢索與分析的一組標準集合，而今日的 metadata 則是隨著 internet 資源的擴展，才被廣泛地運用。

## 二、metadata 的意義及其種類

metadata 一般最普遍的定義是指：「有關資料的資料」data about data」即資料的描述性資訊，從這個定義上來解析，傳統圖書館中的卡片式目錄、紙(書)本式目錄及圖書館自動化系統中的 MARC 目錄都屬於 metadata；國內學者陳雪華教授引 Renato Innella 的看法，認為此種解釋，忽略了 metadata 最重要的概念，即其結構性，認為應將 metadata 定義為” structure data about data”，此「結構」二字使得採用 metadata 做組織資訊的方式和全文索引(Full-text indexing)有所區隔，對資料做描述以建立索引的方法，並非始自今日，但 metadata 一詞卻是在網路資訊出現後，才普遍被使用，且多指網路資訊等電子資訊的描述<sup>5</sup>，換言之，metadata 是提供關於電子資源或資料的一種結構化資料，是對電子資源結構化的描述。

目前 metadata 的種類因其發展背景、目的，以及描述處理資源對象的不同，而產生各種電子資源描述格式(諸如 Dublin Core Element Set、Federal Geographic Data Committee、Categories for the Description of Works of Art、Encoding Archival Description、Text Encoding Initiative、Core Categories for Visual Resources……等等)，其中最受到重視的是「都柏林核心元素集」(Dublin Core Element Set，簡稱都柏林核心集，DC)，中文 metadata 交換格式(MICI)就是以 DC 十五個基本欄位為其主要架構，其十五基本欄位說明如下：

- (一)主題與關鍵詞(Subject)：主要概念的關鍵詞，或其他背景資源等主要詞彙。
- (二)題名(Title)：資源的名稱。
- (三)著者(Creator)：可能是一位或一位以上的個人或機構。
- (四)簡述(Description)：資源的摘要或影像資源的內容描述。

<sup>4</sup> 李郎達，〈metadata 初探〉，吉林《情報科學》第 19 卷第 6 期，2001.63 頁 605。

<sup>5</sup> 陳雪華等《數位圖書館 XML/metadata 管理系統》，台北文華圖書管理資訊公司，2001 年 6 月初版，頁 18。

- (五)出版者(publisher)：發行的個人、機構或政府機關等，發行者與其他重要代理商也可著錄於此。
- (六)其他貢獻者(Contribution)：除著者外，其資源形成有貢獻的其他相關人士或機構(如編主或譯者)
- (七)日期(Date)：其格式一般採 YYYY-MM-DD，此日期不一定與 coverage 字段所定義資源內容所函蓋的日期或期間相同。
- (八)資源類型(Type)：資源的類型或所屬的抽象範疇。
- (九)資料格式(Format)：描述存取、呈現或使用此資源時，可能所需的、硬體工具。
- (十)辨識資料(Identifier)：用來有效辨識此資源的文字或號碼，如 URN、URL、ISSN、ISBN 等。
- (十一)來源(Source)：資源的其他衍生來源。
- (十二)語文(Language)：資源本身所使用的語言。
- (十三)關聯(Relation)：與其他資源的關聯，或所屬的系列等。
- (十四)時空涵蓋範圍(Coverage)：資源所涵蓋的時期和地理區域，時間涵蓋範圍是指作品內容的相關時期，而非資料產生的時間；空間涵蓋範圍若是一個實際地區，則著錄地名或坐標(經緯度)。
- (十五)版權管理(Rights)：被描述資源的版權聲明與使用規範。

基本上 DC 確實是一種 metadata 格式，且基於國際的共識，定義了在網際網路的作業環境中，描述資源的基本資訊，由於 DC 的目的是要讓資源的創造者，或出版者自行建立資源的 metadata，格式簡單易用為其一大特點，都柏林核心集的格式雖然簡單，但為了符合使用者的不同需要，仍然維持了相當的展延性(extensibility)及可變性(modifiability)<sup>6</sup>。

大體而言，國內外數位圖書館 metadata 著錄項目都採行都柏林核心元素集的主要原因，係在於其較全面地概括了網際網路資訊資源的主要特徵，如涵蓋了資源的重要檢索構面(如 Title、Creator、Subject 等項)，輔助檢索及關聯檢索面向(Publisher、Contributor、Identifier、Source、Relation 等項)，以及有價值的說明性資料(如 Description、Date、Type、Format、Language、Coverage、Rights 等項)；另就 Dublin Core 的應用方

---

<sup>6</sup> 同註 5 書，頁 21。

式而言，有兩種方式，其一是將著錄數據與著錄對象存在於不同的文件中，利用 URL 等方式聯結，其二是將著錄數據與著錄對象同時包含在同一個文件中，對數位圖書館的實際運作而言，宜採用第一種方式為佳<sup>7</sup>。

### 三、metadata 的結構與編碼語言

對於一個 metadata 格式來說，其是由多層次的結構所組成，大致包括下列三種結構。

(一)內容結構(content structure)：對該 metadata 的構成元素及其定義標準進行描述；metadata 內容結構需要對所採用的元素進行準確的定義和描述，但是，這些數據元素很可能是依據一定的定義標準來選取的，如 MARC 記錄所依據的 ISBD，因此，metadata 內容結構中需要對此進行說明。

(二)句法結構(Syntax Structure)：即定義 metadata 結構以及如何描述這種結構，例如元素的選取規則、元素描述方法(例如 Dublin Core 採用 ISO/IEC 11179 標準)、元素結構描述方法(如 MARC 記錄結構、XML 結構)等。

(三)語義結構(Semantic Structure)：係指定義 metadata 元素的具體描述方法，尤其是定義描述時間所採用的標準，例如 Dublin Core 建議日期元素採用 ISO 8601、資源類型採用 Dublin Core Types；資料格式可採用 MIME、識別號採用 URL、DOI 或 ISBN 等<sup>8</sup>。

Metadata 編碼語言(Encoding Language)指對 metadata 元素和結構進行定義和描述的具體語義規則，常稱為定義描述語言(Definition Description Language)，在 metadata 發展初期，圖書館界常使用自定義的記錄語言(如 MARC)，但隨著 metadata 格式的增加，必須採行一些標準化的定義描述語言來描述 metadata，其中最重要的應是延伸標記語言(Extensible Markup Language, 簡稱 XML)，一個用 XML 標記的資源是由 XML 前言(XML Prolog)與 XML 實例(XML Instance)組成，其中 XML 前言包括 XML 陳述(XML Declaration)與 XML 文獻類型定義(XML Document Type Definition)，而 XML

---

<sup>7</sup> 參越，臧國全，〈虛擬圖書館中元數據研究〉，武漢《圖書館》，2001年第1期，頁24。

<sup>8</sup> 參閱，張曉林〈元素據開發應用的標準化框架〉，成都《圖書館自動化》，2001年第2期，總號第85號，頁10。

實例則是 XML 實際標記的具體資源內容，字符集及是否引用外部語法規則等；XML 文獻類型訂義則具體定義適用於特定類型資源的標識元素集合；對每一個被定義的元素，XML 進一步定義其屬性名、屬性值表和默認值；另外 XML 還可定義有關實體(entity)和記號(Notation)，透過這些定義，XML 可以自定義一個 metadata 格式結構，因此，所有嵌有 XML 解析器(XML Parser)的系統，能利用這個定義解析 metadata 格式，從而瞭解相應的 metadata 數據。

嚴格而論，在實際網路資訊環境中，我們遇到的往往是用不同 metadata 描述多個不同資源系統所組成的開放型資源體系，因此，需要有效的基於分佈式系統的方法，實現跨 metadata 格式和跨系統的檢索方式，此類檢索的實施，主要是基於公共檢索協議，諸如 Z39.50 協議(用於 MARC 系統和 GILS 系統的開放檢索)等<sup>9</sup>。

#### 四、metadata 系統在國內外的應用--代結語

美國對於數位圖書館的推展相當不遺餘力，較著名的計畫包括美國國會圖書館(Library of Congress)於一九九〇年開始推動的「美國回憶計畫」(America Memory)，進行館內文獻、手稿、照片、錄音、影片等典藏品之數位化，以及編輯成歷史變遷、文化傳承的主題產品<sup>10</sup>；至於 metadata 系統的應用，美國 OCLC 開發聯合線上資源目錄(Cooperative Online Resource Catalog, CORC)，此等均係建立在網路環境之下，以 CORC 而言，乃是為網路電子資源創建書目 metadata 和尋找路徑 metadata 的重要網路資源資料庫。對於我國而言，其中諸多 metadata 實際運作在數位資源建構的計畫中，應首推由行政院國家科學委員會於一九九八年八月邀集中央研究院、台灣大學、清華大學等學者建立的「數位博物館」之推動與合作機制<sup>11</sup>。

建立數位圖書館的本質上思考是為實現電子資源的共建與共享，筆者誠盼在本世紀初期，能經由國內外圖書館界(特別是大學圖書館)的共同努力，能建立起當代完美的大學數位圖書館。

---

<sup>9</sup> 同註文，頁 10-11。

<sup>10</sup> 參閱，同註 5 書，頁 8。

<sup>11</sup> 同前註書，頁 3。