



# Metadata 在電子圖書館所扮演的角色

陳雪華 教授

台灣大學圖書資訊學系

## 壹、前言

台灣地區有許多學術研究單位、圖書館、博物館等機構典藏史籍文獻與各種器物，這些為數甚多的珍藏，為避免損壞，常是無法開放讓大眾參觀。有些或者可以開放參觀，但因時空的限制，許多人無法一睹其收藏品的內涵。現在，透過無遠弗屆的網際網路連結，若能夠將這些寶貴的資源呈現於全球資訊網上，既能夠擴大瀏覽的族群，也能夠讓終將損壞的珍貴資源以另一種方式維持其恆久的生命。

目前呈現於全球資訊網的各類資源有價值的並不多，因此如何將現有的珍貴資源展現於網路上，使其具有恆久的生命，開放給更廣泛的使用群是首要的工作。其次必須針對這些資源的特性加以組織與整理，才能有效地檢索、使用。

## 貳、電子圖書館

有人覺得電子圖書館等於 Database，但這兩年來，我們發現電子圖書館遠遠超

過 Database。電子圖書館的目的有下列三種：

1. 文化保存：將珍藏文獻保存起來。
2. 學術研究：容易讓學者透過網路找尋到所須的資料。
3. 教育學習：網路的使用者最多的是一般大眾，因此應該讓這些珍貴素材具有普及教育的價值，供一般大眾使用。

所謂電子圖書館，不僅只要將各類文件或物件數位化放在網路上供人取用而已，還具有以下幾點特質：

1. 電子圖書館應該是實體圖書館或是資訊檢索系統的延伸，在多媒體、分散式以及協力的網路環境中，提供以使用者為導向的資源儲存、搜尋、處理與檢索的功能。
2. 電子圖書館不僅指數位化的資源配合資訊管理工具，它更意指在一個環境能夠整合資源、服務與人等各因素，以支援資料 (Data)、資訊 (Information)、知識 (Knowledge) 的創造、傳遞、使用、



保存等完整的資訊循環週期，其工作方向在於協助知識的產生、分享與利用，並加速循環過程，使知識的成長比以前快速。

3. 以使用者為導向，提供好用的電子圖書館，良好的組織與整理的工作是極必要的。談到資訊組織與整理，圖書館界在這個領域有非常悠久的傳統與歷史，如：編目規則、機讀格式、分類法、標題表、索引摘要。

### 參、Metadata

在網路上使用資料的方式可以從瀏覽(Browsing)和檢索兩個角度來看。就瀏覽而言，一般即主題指南(Subject Information)，好的電子圖書館會提供這樣的導覽，這是屬於逛街式的，讓使用者對整個內容有概括性的認識，但較花時間。就檢索而言，檢索工具(Search Engines)如要使其更好用，則須透過完整性的Metadata。

#### 一、何謂 Metadata

廣義的Metadata包括檢索工具、主題指南、完整性的Metadata。在主題指南方面，圖書館界常用的有分類法、標題法、其他(如Yahoo、蕃薯藤、院系別等)。我們目前所講的或是網路上所提的Metadata是屬於狹義型的，即完整性的Metadata。

在國內對Metadata有幾種不同方式的

翻譯，尚無一定標準，目前較常使用者，如：資源描述格式、詮釋資料、元資料。

Metadata有以下幾種解釋：

1. Data about data
2. Data describes other data
3. Additional information that is necessary for data to be useful (為使額外的資料更好用，Metadata是必要的)。
4. Metadata是對藏品資料屬性的一組描述，目的在促進資料系統中對資料之檢索、管理、與分析。
5. Metadata在傳統圖書館中即以目錄卡片或OPAC中的機讀編目格式的形式，扮演了資料找尋工具的重要角色。
6. Metadata具有傳統目錄之「著錄」功能，目的在使資料的管理維護者及使用者可透過Metadata了解並辨識資料，進而利用和管理資料。

#### 二、Metadata的必要性

有些人覺得既然網路上已充斥著數位化資料，為何不直接使用，而需要Metadata？縱使如此，Metadata還是有其必要性的，以下分幾方面探討：

##### (一)數位化資料 vs. Metadata

1. Metadata通常比藏品本身小很多，在資料找尋的過程中更容易處理。
2. 目前的技術仍很難做到以非文字物件做為搜尋或自動選取的基礎。
3. Metadata可提供不存在於藏品本身



之資訊（如主題、稽核資料等）。

#### (二)全文檢索 vs. Metadata

Metadata 是經由人工判讀將藏品中的重要資訊抽離或標示出來，這個動作賦予 Metadata 相當強之語意，這是目前自動化技術尚無法真正達到的。全文檢索的缺點在於沒有作 Authority Control，因此可能會發生檢索不到或是檢索到非所需的資訊。為此 Metadata 可與全文檢索相輔相成。

### 三、Metadata 的功能

#### (一)從使用者角度來看

1. 定位 (Location)：如何知道所需的資源在那裡，如 Call Number。
2. 探索 (Discovery)：如何找到所需的資源。
3. 文件記錄 (Documentation)：每一筆 Metadata 都非常詳細記錄被數位化的對象，因此縱使無法看到資源本身，但仍可對文件有相當清楚的認識，有時甚至比看到原文還容易了解。
4. 評估 (Evaluation)
5. 選擇 (Selection)

#### (二)從系統角度來看

1. 描述藏品特徵
2. 提供瀏覽及檢索功能
3. 管理功能（如系統識別號）
4. 組合各個物件 (Objects) 以及藏品

的再呈現

### 四、Metadata 的格式

Metadata 的種類目前無可計數，以下介紹幾類較為重要的 Metadata 格式：

#### (一)描述科技文獻

1. BibTeX：LaTeX 的一部份
2. EELS：瑞典技術大學工程電子圖書館
3. EEVL：英國愛丁堡工程虛擬圖書館
4. RFC 1807：透過電子郵件傳遞技術報告

#### (二)描述人文及社會等學科資源

1. ICPSR SGML Codebook Initiative：屬於社會科學 data sets，特別是政治學的資料。
2. TEI (Text Encoding Archival) Headers：用在人文學及語言學上，促進資源的交換及協助人文學者充分利用電子資源。

#### (三)描述政府資訊

GILS (Government Information Locator Service)：由美國聯邦政府訂定，為一種國家標準。

#### (四)描述地理空間性資源

FGDC (Federal Geographic Data Committee Standard)，又稱 CSDGM (Content Standards for Digital Geospatial Metadata)：為美國聯邦政府



所訂定，加州大學聖塔芭芭拉分校 Alexandria Digital Library 就是用此格式來描述。在眾多的 Metadata 中，此格式發展的最為成熟，亦發展出來 Authoring Tool。

#### (五)描述博物館藏品及檔案特藏

1. CDWA (Categories for the Description of Works of Art)
2. CIMI (Computer Interchange of Museum Information)
3. EAD (Encoding Archival Description)：描述檔案的格式，提倡者為 LC (Library of Congress)，亦深受 MARC 影響。

#### (六)描述大量網路資源

1. URCS (Uniform Resource Characteristics/Citations)：與 URI、URL 同一機構制定。
2. Dublin Core (Dublin Metadata Core Element Set)：由 OCLC 推展出來。

### 五、Metadata 的來源

Metadata 的創作者與來源可分為作者、資源儲存的管理者、其它單位的創作者。針對電子圖書館而言，其 Metadata 創作者大都屬於資源儲存管理者。

### 六、Metadata 的著錄層次

- (一)無結構化的索引，例如 Search Engine 自動擷取的資料。

(二)相當結構化，但較精簡，提供足夠資源描述資訊，例如 OCLC 的 NetFirst、Dublin Core。

(三)相當結構化，並且非常複雜，有其特殊領域，提供詳細的資源描述資訊。

### 七、Metadata 實施現況

Metadata 在實施應用方面，除 MARC 外，其他均處於發展階段。MARC 因其記錄量多、架構詳細、且有編目規則作為描述的基礎，使其在描述有形館藏時技術成熟且歷史悠久；然而傳統的 MARC 應用於目前的網路環境稍嫌不足，因此從 1992 年起，即針對 MARC 作了一些修改：

(一)欄號 856：用以描述電子資源，可 Hyperlink 電子資源本身。

(二)欄號 505：將目次內容掃描呈現（主要對象還是紙本式資料）。

### 肆、Dublin Core

Dublin Core 於 1995 年 3 月 OCLC/NCSA Metadata Workshop 會議中被提出（第一次 Dublin Core 研討會）。

#### 一、何謂 Dublin Core

它是一組描述網路資源之簡單欄位，提出時只有 13 個基本欄位。Dublin Core（簡稱 DC）的描述對象為文獻式資源 Document-Like-Objects。

#### 二、Dublin Core 的基本原則

(一)簡單



在格式及欄位名稱上力求簡單，使用簡便，不需專業訓練。其優點如下：

1. 作者可自行建立 Metadata；2. 軟體廠商可自行發展編輯器 (Authoring Tool)，並可延展；3. 可作為進一步書目描述的基礎；4. 不同群體間語意的共通性。

### (二) 彈性

DC 有四大特性：延展性 (Extensibility)、選擇性 (Optionality)、可重覆性 (Repeatability)、可變性 (Modifiability)，具有相當彈性以符合不同群體的需求，且可利用 Qualifiers 如 ("Sc-heme", "Type")，對外連結到更完整的紀錄。

DC 的利基點為改進目前 Search Engine 以沒有結構的關鍵字描述之缺點，且 DC 簡單的格式、結構化描述、及所具有的延展性，使其在未來可再進一步的深入描述。

### 三、Dublin Core 的發展

DC 的初期應用起源於用來描述網路上數量最龐大的“文獻式資源” HTML 文件，但未標準化。

DC 的第一次研討會有些未解決的問題：1. 彈性與一致性的衝突。2. 沒有清楚的結構及用法。3. 沒有標準的應用方法。

第二次的 DC 研討會於 1996 年 4 月在英國的 Warwick 舉行，會中最重要的是

提出了 "Warwick Framework" 的觀點。所謂 Warwick Framework 是一個可將不同形態的 Metadata 物件分別包裝起來，以進行資訊交換的架構，它必須具有能將不同的 Metadata 物件視為單獨的模組、同時須考慮到未來出現新的 Metadata 格式、及內部的 Metadata 物件可以連結到外部等特性；利用 Framework 可使眾多的 Metadata 語意共通，讓不同 Metadata 間可以相互對映。另外還提出使用指南，由作者、資料保存管理者製作。

至於 DC 應用在 HTML 文件方面，即 DC 與 HTML 的對映，1996 年 W3 Conference 會中提出將 Metadata 融入 HTML 文件的建議。

第三次 DC 研討會於 1996 年 9 月在 OCLC 總部召開 (CNI/OCLC Image Metadata Workshop)，將影像資料視為 "Document-Like-Objects" 是此次會議的最大收穫；並將原來的 13 個欄位稍加修改，增加 Description、Rights Management 兩個欄位，擴張成 15 個，即可用來描述影像資料。此 15 個欄位如下：

1. 標題與關鍵詞 (Subject & Keywords)
2. 題名 (Title)
3. 作者 (Author & Creator)
4. 描述語 (Description)
5. 出版者 (Publisher)
6. 其他貢獻者 (Other contributors)



7. 日期 (Date)
8. 資源的類型 (Resource type)
9. 資料格式 (Format)
10. 資源識別碼 (Resource identifier)
11. 關係 (Relation)
12. 來源 (Source)
13. 語言 (Language)
14. 涵蓋範圍 (Coverage)
15. 權限管理 (Rights management)

DC 在影像描述上的應用，以 CIMI (Consortium for Computer Interchange of Museum Information，美國博物館資訊交換協會) 為例，即是採用 DC 來描述影像資料。

第四次 DC 研討會在澳洲坎培拉舉行，此時 DC 漸趨成熟，發展階段結束，規劃佈署階段開始。此時有三個重點：欄位組及 Qualifiers 的正式確定；延展性的爭議；某些欄位必須更清楚的定義以及 Canberra Qualifiers 的產生。

在延展性的爭議方面，如要維持 DC 簡單及一致的特性，則須有條件地使用 Qualifiers，但將會犧牲掉很多藏品的特質。但盡量發展 Qualifiers 來滿足不同群體的不同需求則會與 DC 要求簡單、一致的特性相衝突。因此最後決議，為使其可以延展、也可以一致，而有 "Canberra Qualifiers" 的產生。

Canberra Qualifiers 有三類：Lan-

guage 描述時所用的文字（非所描述的資源所用的文字）、Scheme、Type。

Qualifiers 使用實例：（採用 HTML 第二版）

1. <META NAME="DC.Title" content=" (LANG=en) Cataloging of Internet resources">
2. <META NAME="DC.Creator" content=" (TYPE=editor) Hsueh-Hua Chen">
3. <META NAME="DC.Subject" content=" (SCHEME=LCSH) Cataloging of computer files">
4. <META NAME="DC.Identifier" content ="(SCHEME=ISBN) 0-670-77289-5">

第五次 DC 研討會於 1997 年 10 月在芬蘭赫爾辛基舉行，會中達成以下幾點共識：

1. 加快標準化的腳步。
2. 區分簡單、複雜兩種 Dublin Core 格式：即有無使用 Qualifier。
3. 語法上以 HTML 和 RDF 格式為主。
4. 針對未有定論的議題成立工作小組（如 Date、Relation、Rights Management）。
5. 次欄位（項目）及 Qualifiers 使用的制定原則。

Dublin Core 的價值在於：

1. 激勵作者與出版者提供這些詮釋資料的項目。



2. 激勵網路出版工具中加上產生詮釋資料的樣板 (template) 在內。
3. 激勵軟體廠商發展 Authoring Tool , 使資料容易鍵入。
4. 可做為詳細編目之基礎。
5. 記錄容易被了解。

#### 四、Dublin Core 的應用實例

Dublin Core 從 1995 年發展至今，時間雖不長但使用者衆。其運用實例如下：

1. The Nordic Metadata Project
2. DSTC (Distributed Systems Technology Center) — TURNIP、HotOIL、MetaWeb
3. OCLC Office of Research
4. The National Document and Information Service
5. Mapping between the Dublin Core and MARC records
6. 加州大學聖塔芭芭拉分校 Alexandria Project
7. 加州柏克萊大學 Digital Library 計畫
8. CIMI (Consortium for Computer Interchange of Museum Information)

#### 伍、以台灣大學電子圖書館為例

台灣大學電子圖書館 / 博物館 Metadata 研究小組於民國 86 年 3 月成立。其系統組成部份分為：Metadata、資料 (Data)、架構 (Framework)。

#### 一、工作目標

制定 Metadata 以及權威檔 (Authority Files)；權威檔又分為名稱權威檔 (人名、地名等)、及主題權威檔 (索引典) (由臺灣大學圖書資訊學系黃慕萱教授負責此部份工作)。

在制定 Metadata 方面，工作目標如下：

1. 符合藏品特性。
2. 提供使用者必要的檢索點，符合使用者資訊尋求行爲。
3. 考慮與其他電子圖書館資料交換上的共通性，發展可共同使用的 Metadata。
4. 品質 (精緻編目與簡略編目) 與成本之考量：由於所編目的資料很珍貴，因此詳細編目是有必要的。

名稱權威檔的功能在於辨識並追蹤名稱的各種更替情形，維護資料庫檢索項 (Access Point) 文辭形式的一致性，及協助資訊檢索。

主題權威檔的功能則在於將相關主題之資料聚集一處，協助資訊檢索；展現本電子圖書館主題詞彙之間的層屬關係，幫助了解其知識架構並擴大或縮小檢索範圍；提供標準辭彙以確保資料索引的一致性，可以解決同義詞、同形異義詞的問題；採取控制辭彙，有良好的結構，使用時可選擇最適當之用語，以提高檢索的精確率 (Precision Ratio) 和回現率 (Recall)



Ratio)。

## 二、工作項目

1. 探討藏品之背景與特性。
2. 研究國內外現有之 Metadata 格式，但無法全部使用。
3. 探討 Metadata 與資料庫及整個系統之關係。
4. 調查潛在使用者的資訊需求與搜尋行爲（由台灣大學圖書資訊學系林珊如副教授負責此部份工作）。

## 三、目前的工作進度

1. 調查現有 Metadata。
2. 探討使用現有 Metadata 之可行性。
3. 制定古文書部份的 Metadata（暫行性）：淡新檔案、岸里大社文書、伊能文庫、人類學系古契書。
4. 制定淡新檔案行政篇的索引典。

## 四、未來的發展

1. 古文書 Metadata 格式之繼續改良。
2. 其他藏品種類（器物、地圖與其他類資料）Metadata 的格式發展與設計。
3. 現有國內外幾個重要 Metadata 格式發展之觀察與追蹤（如 CIMI、Dublin Core、EAD）。
4. 探討資訊交換的可行性，如對映 (Mapping) 自己的 Metadata 到一個標準的 Metadata（可能考慮用 Dublin Core）。讓其不只是作國內數位圖書館的交換，亦可跟國際作連結。

5. 發展完整的索引典。

6. 制定人名權威檔，未來發展其他權威檔（如地名）等。
7. 調查中央研究院史語所、國家圖書館、國家圖書館台灣分館、故宮博物院、台灣省文獻會等單位描述其藏品（檔案、文獻、器物等）的方式。
8. 建議適用於描述上述之珍藏電子文物（獻）之 Metadata 與資訊組織模式，作為設計中文環境的電子圖書館系統的基礎。

## 陸、國科會「數位博物館專案計畫」簡介

目前這個計畫正在進行中。其核心參與單位為中研院與台灣大學。此計畫起源於國科會想將國內較好的博物館資訊數位化，供一般大眾使用。台灣大學電子圖書館研究小組則針對此計畫提出「數位博物館專案計畫」資源組織與檢索之規範。計畫的研究範疇為在中文環境中數位圖書館/博物館有關於資訊組織與檢索標準各個層屬的重要議題。

根據文獻探討與歸納過去心得，我們認為至少有以下五個層面的議題，需在本計劃中加以探討：

1. 電子資源之組織與中文資源著錄格式的制定。
2. 使用者資訊尋求行爲分析與便易使用環境之建立。





3. 索引典架構及權威資料庫的編纂。
4. 資訊檢索及資料搜尋服務系統的設計。
5. 數位博物館 / 圖書館整合檢索 ( 現在用 Z39.50 ) 。

此外，雖然目前是中文資料，但為了方便與國際間作交換，多語言的轉換亦需探討。

#### 一、計劃內容

本計劃主要支援台灣大學 ( 淡水河溯源 )、暨南大學 / 科博館 ( 蝴蝶生態面面觀 )、中研院 ( 中華文化、臺灣文化 ) 之雛形系統。未來研究重點在於制定國內之中文數位圖書館 / 博物館環境中有關資源組織與檢索之相關標準。這些標準須要注意與國際相關標準能夠相容的問題。參與國際性組織 ( 如： **Consortium for the Computer Interchange of Museum Information, CIMI** )，是促進國內數位博物館 / 圖書館系統國際化的重要管道。

#### 二、計劃項目

1. 總計劃：數位博物館專案計劃，資源組織與檢索之規範。
2. 子計劃一：電子資源之組織與中文資源著錄格式的制訂。
3. 子計劃二：使用者資訊尋求行為分析與便易使用環境之建立。
4. 子計劃三：索引典架構及權威資料庫之建置。
5. 子計劃四：資訊檢索及資料搜尋服務系統的設計。
6. 子計劃五：數位博物館 / 圖書館整合檢索。

#### 柒、結語

1. 電子資訊的威力無所不在，取之不盡、用之不絕；可廉價快速的傳播，促進知識的成長，是表達知識的通用媒介。
2. 電子圖書館 / 博物館的重要性漸增，我國須發展具有特色的中文電子圖書館 / 博物館。
3. 電子資訊組織與整理的重要。

\*\*\*\*\*  
 \* 本文為「二十一世紀大學圖書館的演變與發展研討會」演講記錄，由陳鳳玲小 \*  
 \* 姐記錄，並經主講者寓目同意刊登。 \*  
 \*\*\*\*\*