

台大電子圖書館與博物館數位化 資源之組織

陳雪華

國立臺灣大學圖書資訊學系教授兼系主任、所長

張玉欣

研究助理

The Role of Metadata in National Taiwan University Digital Library/Museum Project

Hsueh-Hua Chen Yu-Shin Chang

Abstract

National Taiwan University (NTU) is one of the oldest universities in Taiwan and hence possesses a lot of rare books, historical texts, and precious documents. However, the access of these collections seems unfortunately to be limited only to the minority of related researchers inside the campus. This is against the overwhelming trend of information resources sharing. The best way to render these collections available to users all over the world in a more efficient way is to distribute these valuable collections via the powerful WWW.

To achieve this aim, NTU Digital Library/Museum (DL/M) Group needs to digitalize these collections and establish metadata for them. This article describes the importance of metadata in digital library, gives a brief introduction to the development of NTU DL/M project, the system architecture of metadata in NTU DL/M project, and finally raises some problems which need to be solved as well as the future development of metadata in NTU DL/M project.

關鍵詞：數位化資源、電子圖書館、電子博物館、

Matadata、Digital library、Digital museum

壹、前言

1975年，Roger W. Christian出版"The Electronic Library"一書，首先提出「電子圖書館」的概念。❶另外，美國圖書館學大師F. W. Lancaster在1978年透過其著作“Toward Paperless Information Systems”，預言在公元二千年後電子出版品將漸替代紙本式資料，未來的圖書館將可能進入以電子媒體為主要館藏的「電子圖書館」（Electronic Library）時代，因此這一名詞在圖書館界造成很大的震撼，且沿用了二十年之久。❷

近年來，由於各種數位化技術相繼興起，於是「數位化圖書館」（Digital Library）漸成為新興名詞。尤其是1992年前美國參議員高爾（Al Gore）（目前為美國副總統）提出「資訊基礎建設與科技法案」（The Information Infrastructure and Technology Act of 1992），自此之後，有關數位化圖書館之各項研究計劃、國際會議與研討會等便陸續展開，不只受圖書館界人士的矚目，更是在資訊、電腦界中廣受熱烈討論。❸可以預測數位化圖書館在二十一世紀中將扮演一個相當重要且具影響力的角色。但由於「電子圖書館」一詞在台灣較為響亮與普遍，因此以下均沿用電子圖書館之名稱。

所謂電子圖書館，不僅只是將各類文件或物件數位化放在網路上供人取用而已。在全球資訊網中，資訊過於氾濫、無序，常常無法有效地找到資訊。電子圖書館應該是從實體圖書館的延伸或是從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之儲存、搜尋、處理與檢索的功能。❹換言之，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。❺

從另一個角度來看，電子圖書館實為一科際整合之全新領域。它不僅

結合了圖書館學、電腦與資訊工程、以及與電子圖書館之內容相關的學科等早已存在的學術領域，許多新的研究議題，如：資料的數位化相關研究，電子資訊描述、組織、與交換，新的電子資料之儲存與擷取方法等研究也隨之快速發展。

在電子圖書館的環境中，要有效地提供資訊的利用，良好組織、整理的工作是極必要的。圖書館界對於圖書資訊的組織與整理有悠久的傳統，如編目規則、機讀格式、分類法、標準表以及索引摘要等方式，因此圖書館界一向自詡為此領域的專家。然而，目前在網路環境中，對於資訊的組織整理有興趣者，除了圖書館界人士之外，還有電腦與資訊工程界或是各學科領域的專家們，紛紛在探討此相關議題，並制定各種metadata以因應資訊檢索之需。

Metadata是敘述一個資源屬性的資料，通常稱為“Data about data”或是“Data describes other data”^⑥、“Additional information that is necessary for data to be useful”^⑦“information about the data that helps in optimization and management of that data”^⑧等廣狹不一的定義。其翻譯名詞亦眾多，包括「詮釋資料」、「資源描述格式」、「超資料」或是「元資料」等^⑨。

本文的重點將以「台大電子圖書館與博物館」計畫為例，探討電子圖書館中資料之描述與組織的問題。首先將探討metadata之重要性，其次再對本計畫作一概略的介紹，並說明metadata在本計畫中所扮演的角色，以及系統架構中的metadata處理，最後說明一些尚待解決的問題並探討未來發展的方向。

貳、Metadata在電子圖書館的重要性

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。^⑩它在傳統圖書館中即以目錄卡片或線上公用目錄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，其他如索引或摘要等另一種形式之metadata，也是不可或缺的工具。David Levy

認為編目 (cataloging) 是透過一組有系統有組織之「代替物」 (surrogates)，將圖書館館藏排列成序並且提供檢索的過程。①編目提供了館藏之屬性，例如：題名、作者、主題等；有些屬性並不真的存在於所描述之對象本身，例如：分類號、標題等。這些屬性就成為圖書館排列、組織、與檢索館藏的根據。

在電子圖書館中，硬體之儲存容量與運算能力已不再是問題，而數位化之後的藏品本身則具備了電腦運算之價值，全文輸入讓全文檢索成為可能就是一個很好的例子。在這樣的情況下，編目及「代替物」（指 metadata）作為資料發現之工具有其存在的必要嗎？有人認為資料之數位化形式既已存在，就可捨“代替物”而以資料本身作為資料發現之基礎。Lynch、Michelson等人則提出以下四點作為反駁：②

1. 「代替物」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. 在智慧產權的考量上，資訊提供者可能較願意提供「代替物」，而非直接提供藏品本身。
3. 目前的技術仍無法做到以非文字 (non-textual) 物件做為搜尋或自動選取之基礎。
4. 「代替物」可提供不存在於藏品本身之資訊。(例如主題、稽核資料等)

除了以上四點，metadata也具有傳統目錄之「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，進而去利用和管理資料。這個功能，在資訊爆炸的網路環境中更形重要。

另外，由於metadata是經由人工（館員或專家）判讀，將藏品中的重要資訊抽離或標示出來，這個動作賦予了metadata之元素相當強之語意 (semantics)，這是目前自動化技術尚無法真正達到的。例如：系統可直接分辨「人名」一欄中的值代表的就是一個人名，而非只是一個無意義的字串，這對系統建立物件間的關係有相當大的幫助。由此可知 metadata在電子圖書館中占有相當重要的地位。

目前存在的metadata形式繁多，很多電子圖書館由於其藏品性質特殊

而發展適合其藏品特性之資料描述格式，例如：描述政府出版品的GILS (Government Information Locator Service)，及描述地理空間資料的FGDC (Federal Geographic Data Committee Standard)。有些則繼續採用成熟的機讀編目格式，這在資訊交換上有相當的優勢。另外如最近引起廣泛注意的Dublin Core亦被視為有潛力之電子資料描述與交換格式之一。

叁、Metadata與台大電子圖書館與博物館

一、台大電子圖書館與博物館計畫簡介

台大電子圖書館與博物館計畫 (<http://ntudlm.ntu.edu.tw>) 的目的是要建立一個以瀏覽器為使用方式的電子圖書館。計畫的內容主要為台灣史料，由於這些藏品年代久遠非常脆弱，需要特殊的保存環境，少有機會能展現其風采，本計畫即在將這些珍貴的文化遺產以數位化的方式保存、呈現、並提供給大眾利用。本計畫是跨文學院與工學院的大型群體合作計劃，研究人員的專業領域涵括圖書館學、資訊工程、歷史學、人類學等，集合了人文與科技專業領域的專家。本計畫初期數位化的對象為台灣大學所珍藏的清中葉以降台灣古文書，目前正進行《淡新檔案》與《台大人類系藏古契書》的數位化。計畫之目標為：

1. 提供歷史與人類學者一個完整的研究環境
2. 提供一般大眾有關台灣歷史文化的教育媒體
3. 整合台灣大學所收藏的台灣珍貴史料、藏品、與研究成果
4. 透過數位化保存台灣本土文化遺產
5. 推動中文電子圖書館之研究

電子圖書館之研究為一新興領域，國內之圖書館學與資訊科學界在這個領域可說是仍處於摸索階段。本計畫之研究群在一年來的不斷學習與摸索中，漸漸歸納出以下研究議題：

1. 電子圖書館文獻與博物館藏品數位化技術與應用之研究
2. 電子圖書館與博物館資料儲存系統之設計

3. 台灣史料中知識擷取與覽讀環境之研究
4. 電子圖書館與博物館中影像處理
5. 台灣史料整理與數位化
6. 台灣史料數位化資源組織 (metadata) 之研究
7. 電子圖書館索引典 (thesaurus) 建構之研究
8. 台灣史料使用者資訊需求與搜尋行為之研究
9. 檢索詞彙/索引詞彙對映之研究
10. 分散式電子圖書館與博物館系統之整合檢索研究

這些研究議題所牽涉之知識範疇跨越人文與科學，而每一個議題相互之間又環環相扣，缺一不可。舉例而言，**metadata**之研究成果可以提供“電子圖書館與博物館資料管理及儲存系統設計”做為儲存並管理時的基礎，亦即每一筆資料在輸入之前必須先依**metadata**格式來描述並組織資訊。由於**metadata**所描述的資訊可供檢索之需，因此與“分散式電子圖書館與博物館系統之整合檢索研究”有極大的相關性。此外，**metadata**欄位之設計必須考量使用者的檢索需求，“台灣史料使用者資訊需求與搜尋行為之研究”之研究成果可提供本研究之參考並修正。

二、Metadata在台大電子圖書館與博物館中之功能與角色

在說明**metadata**在台大電子圖書館與博物館之角色與功能之前，必須先介紹台大電子圖書館與博物館之系統結構。此系統在概念上是由三大部份所組成：

1. **Metadata**：每一件藏品都有一筆描述它的書目記錄。目前已數位化的藏品類型主要為古文書，由於其性質與書籍大不相同，在描述單位上不似一筆書目記錄即是描述「一本書」如此單純，因此必須對什麼是「一件藏品」加以定義。例如：淡新檔案共有1,163「案」，每案之下又有數量不等的「件」，共有19,152件，處理的原則是以藏品對使用者來說有意義之最小單位為描述之對象。
2. **資料 (data)**：即數位化的資料本身。目前有全文及影像兩種，在未來可能會增加聲音、影片等多媒體型態資料。
3. **架構 (framework)**：為以瀏覽器呈現數位化藏品時的網頁及其所

構成的架構。

此三個部份之關係非常密切，例如每一件藏品資料（data）都連結著一筆描述著其屬性之metadata記錄，使用者透過瀏覽或檢索的過程，找出所需要的藏品資料之metadata記錄，並透過這些metadata將藏品資料從儲存庫中調出，組成架構（framework）呈現給終端機前的使用者。當然，這個過程絕大部份是透明的，使用者可能根本感覺不到，更不會完整地看到這些metadata記錄。

數位化工作規劃開始時，在資料之組織與整理工作上，研究者首先遇到的挑戰就是必須為所要數位化的藏品選擇或設計適合的metadata描述格式。為了解決這個問題，特別組成了「Metadata 研究小組」，立即開始著手相關之研究。工作項目包括瞭解藏品之背景與特性、研究國內外現有之metadata格式、瞭解metadata與資料庫及整個系統之關係、以及潛在使用者的資訊需求與搜尋行為。研究小組認為，metadata格式的考量重點為符合藏品特性、提供使用者必要的檢索點、未來與其他電子圖書館資料交換上的共通性、以及品質與成本之考量。

本計畫在現階段數位化的藏品型態主要為古文書，包括《淡新檔案》、《台大人類學系藏古文書》等。在瞭解古文書的性質之後，研究小組深入研究許多描述類似藏品之metadata，包括描述博物館藝術藏品的CIMI（Consortium for Computer Interchange of Museum Information）及描述檔案資料的EAD（Encoding Archival Description），然而由於文化及藏品性質的差異，這些 metadata無法完全適用於中文特藏。

此外，基於交換上的共通性之考量，研究小組亦曾考慮採用機讀編目格式，因其為發展成熟之交換格式，但在評估其適用性之後，發現機讀編目格式對於古文書藏品來說過於複雜，且其原來描述書籍為主的特性無法滿足我們特殊藏品的需求，舉例來說，在古契書中「作者」的概念並不明顯，但「關係人」則為重要之檢索點之一，若要勉強將這些資訊塞入機讀格式中具類似意義但不完全相同的欄位，則會造成語意（semantics）上的損失，這是研究小組所不願意見到的。再者，處理機讀編目格式需要特殊且複雜之軟體，將成為系統設計上不必要的負擔，因此在成本與效

益的考量之後，決定自行設計一個適合中文特殊藏品之metadata，但是設計上仍保留了許多機讀編目格式的精髓。

在設計古文書metadata格式之過程中，研究小組成員不斷地與學科專家、資料使用者、使用者資訊尋求行為研究專家、與系統設計人員討論與溝通，經過許多的努力，台大電子圖書館與博物館之古文書metadata格式（暫行性）終於成型（請參見附件）。

一般而言，metadata所提供的功能包括：1.定位（location）：如何知道所需資源的儲存位置所在；2.探索（discovery）：如何找到所需的資源；3.文件記錄（documentation）：描述並記錄文件的性質與內涵；4.評估（evaluation）：協助使用者判斷資源對其之價值；5.選擇（selection）：幫助使用者決定是否取用該資源。台大電子圖書館與博物館系統中的metadata亦提供上述之各項功能，另外，從其他角度來說明metadata在本系統中的功能如下：

● 描述藏品特徵

一如傳統圖書館中的一張目錄卡片，或是線上公用目錄中的一筆機讀格式目錄，它所描述的是某一件館藏的特徵，在台大電子圖書館與博物館系統中的一筆metadata記錄，描述的是某一件藏品原件之屬性與特徵。在描述格式設計的過程中，除了資訊描述專家之外，使用者行為專家、與學科專家（歷史學者與人類學者）亦積極參與討論，務必使藏品每一個重要特徵及屬性都被充份地描述出來，並建立檢索點。

● 提供瀏覽及檢索功能

正如Levy所認為的，在將藏品之特徵描述出來之後，可利用這些特徵或屬性將藏品以某種方式組織起來，並加以排序。● 台大電子圖書館與博物館提供了兩種檢索資料的方式：

a. 瀏覽：利用資料的某些屬性排列藏品，符合使用者所熟悉的“逛書架”式的行為，在電子環境中提供藏品之架構，可以對藏品全貌有概括性的認識，避免在電子環境中喪失“方向感”。

b. 搜尋：metadata提供了搜尋之基礎。在著錄的過程中，將資料中的重要資訊抽出加以組織，賦予語義，並建立關係，將使檢索結果更加精確

。

●管理功能

Metadata中之管理性欄位支援系統管理之功能，例如：系統識別號。

●組合各個物件 (objects) 以及藏品的再呈現

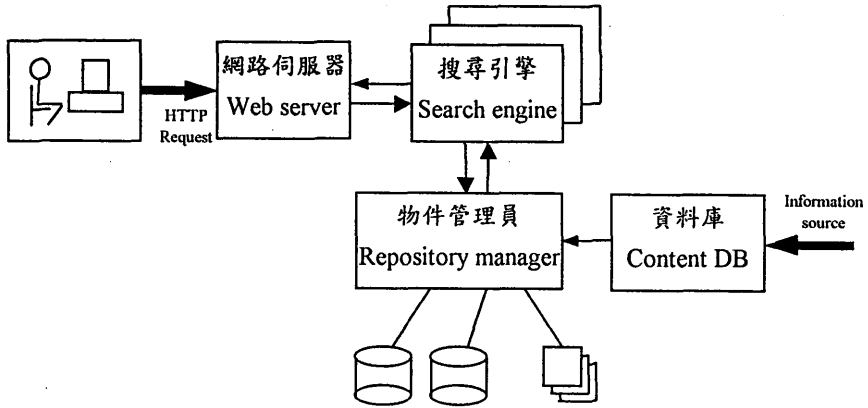
藏品數位化之後，是以物件 (objects) 的形式散落放置，一個物件可以是一幅影像、或一篇全文等。以一件古契書為例，其數位化影像為一物件，而其全文檔為另一物件，二者在物件儲存庫中分別獨立存放，而 **metadata** 中則記載了這些物件的 **handle**。當使用者搜尋到此一古契書時，系統就需要找出屬於這件古契書的各個物件，加以組合，才能呈現給使用者一件完整的藏品。

本研究小組並不以建構一個自己可使用的電子圖書館為滿足，惟有加強 **metadata** 之共通性 (interoperability) 及朝向開放式系統的設計，才能提供共通且整合性之檢索。本計畫一直進行網際網路資訊交換協定主從架構之相關研究，由於系統架構本身仍未十分成熟，因此目前尚未進入系統架構中的實作階段。也因為如此，在設計 **metadata** 時保留了相當的彈性，當未來系統臻於成熟，資訊交換協定之主從架構進入實作階段時，可以很容易的將現有的 **metadata** 轉換成將採用的標準交換格式。

三、台大電子圖書館與博物館系統架構中之 **metadata**

為了對台大電子圖書館與博物館中的 **metadata** 之特性有較完整的認識，以下概略說明 **metadata** 記錄在進入之後，系統處理的情形。

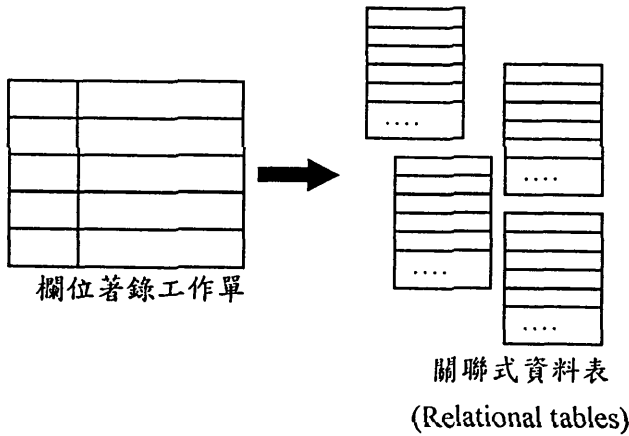
台大電子圖書館與博物館之 **metadata** 具有相當大之彈性，這與系統架構中對 **metadata** 的處理有關。圖一為台大電子圖書館與博物館系統架構圖，**metadata** 記錄在進入系統之後是與其他數位化資料一起被放置在資料管理員 (content manager) 中，接下來資料管理員中的資料便會被複製一份，轉換成一個個 "物件"，並存放在物件儲存庫 (object-based repository) 中。



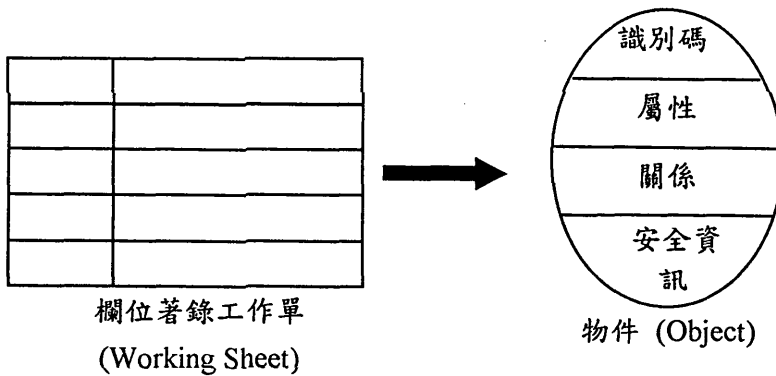
圖一：台大電子圖書館與博物館系統架構圖

從附件[台大電子圖書館與博物館古文書Metadata欄位簡表]中可看出 metadata資料結構的複雜性，如果完全以關聯式的概念來處理，則必須面對 metadata格式在以資料表為基礎的關聯式資料庫中被拆開為數個部份來儲存的問題。(參見圖二)但是一筆 metadata就應該被視為一個完整的記錄，若系統在運作時必須隨時組織 metadata記錄，將造成系統運作上效率低落的問題。

基於以上的理由，研究小組決定以 "物件" 的觀念來處理系統架構中的 metadata。在這裏必須先對 "物件" 下定義：所謂 "物件" (object) 就是最基本的資訊單位，換句話說，一筆 metadata記錄或一幅影像都可視為一個物件。在台大電子圖書館與博物館系統中，一個物件包含三個部份：
1. 識別碼 (Identifier)：物件在系統中獨一無二之識別碼。
2. 安全資訊描述 (Security information description)：安全資訊描述係對使用者權限之控制。
3. 屬性與關係 (Attributes and relations)：原來 metadata欄位中的資訊在此成為一個個的屬性並建立關係。(參見圖三)



圖二：一筆metadata記錄被拆成數個關聯式資料表。



圖三：一筆metadata記錄被包裝成物件。

以物件的方式處理metadata的優點是將單一的資訊單位包裝成單一之物件，而非分散的多個關聯式表格資料，所有的屬性與關係皆被包含在物件內。此外，metadata物件是獨立於metadata格式之外的，因此在未來進行資料交換時，無論所選擇的是那一種資料交換格式，系統都能輕鬆地將

metadata物件轉換成交換格式。以物件為基礎的另一個好處是未來在多媒體資料之處理上，多媒體物件模型將會占相當之優勢。

肆、問題與未來發展

雖然本metadata研究群的努力有一初步的成果，但是這僅是古文書類藏品暫行使用的metadata而已。尚有許多問題有待解決，例如需要進行大規模使用者資訊需求研究，另外，此metadata格式亦需要與其他標準的metadata相對映 (mapping)，以便與他人資料交換。隨著計畫繼續進行，仍有許多工作必須持續，未來工作方向為：

- 古文書metadata格式之繼續改良
- 繼續探討現有國內外幾個重要metadata格式發展之觀察與追蹤
- 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 (可能為Dublin Core)
- 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
- 發展索引典
- 制定人名權威檔，未來發展其他權威檔 (如：地名) 等

未來，本研究小組期望所規劃設計之metadata，不僅可以符合台大電子圖書館與博物館藏品的特性，並符合使用者資訊尋求行為，還可以與其他電子圖書館metadata相容。更期望所設計的metadata欄位著錄工作單，可提供其他相關單位 (如：中研院、台灣省文獻會等) 實際輸入資料之用，可解決與其他電子圖書館資料交換之問題。

附件

台大電子圖書館與博物館古文書Metadata欄位簡表

1997年7月8日 本計畫Metadata工作小組訂定

	欄位名稱	次欄位名稱	淡新檔案	人類學系古文書
1	記錄識別號		★	★
2	典藏排架號	\$a排架號	★	
		\$b館藏地點		
3	收藏單位登錄號		★	★
4	藏品類型		★	★
5	來源		★	★
6	文書形式類別	公文書(再細分)	★	★
		私文書(再細分)		
7	集叢名		★	
8	題名		★	★
9	時間		★	★
10	發文者		★	★
11	主要收文者(R)		★	
12	次要收文者(R)		★	
13	主要關係人(R)	\$a人名(名稱)	★	★
		\$b身份		
		\$c文書角色		
		\$d族名		
		\$e社名		
14	次要關係人(R)	\$a人名(名稱)	★	★
		\$b身份		
		\$c文書角色		
		\$c族名		
		\$d社名		
15	地名(R)	\$a省	★	★
		\$b府、州		
		\$b縣、廳		
		\$c堡、里		
		\$d莊、街、社		
		\$e土名		
		\$f現今地區代碼		
\$e其他				
16	內容性質分類(R)		★	★
17	內容摘要		★	★
18	一般附註(R)		★	★

	欄位名稱	次欄位名稱	淡新檔案	人類學系古文書
19	相關參照註		★	★
20	來源註	\$a 入藏日期		★
		\$b 取得方式		
		\$c 捐贈出售者		
		\$d 採集者		
		\$e 採集地點		
21	相關族群註		★	★
22	其他媒體註			★
23	稽核資料	\$a 頁數	★	★
		\$b 高廣尺寸(長×寬)		
		\$c 附件		
		\$e 插圖		
24	印記	章:職章	★	★
		印:官印		
		手、腳模		
25	編目員		★	★
26	編目日期	紀錄建立日期	★	★
27	檔案大小		★	★
28	影像檔品質		★	★

● (R) 為可重複使用欄位。

附 註

- 胡述兆總編輯，**圖書館學與資訊科學大辭典**。初版。台北市：漢美（民國84年），頁1819。
- 張莉慧，**數位化圖書館在博物館之應用**，台中縣東勢鎮：國寶魚（民國85年），頁24。
- Digital Resources Organization and Projects <<http://www.blc-bnc.ca/ifla/diglib.html>>
- Maria Aemankova, "Workshop Report : Distributed Knowledge Work Environments" <DIGLIB@INFOSERV.NLC-BNC.CA>
- 同上註。
- Terry Kuny, <Terry.Kuny@xist.com> "Metadata : What is It?" 24

- Apr.1997, <DIGLIB@INFOSERV.NLC-BNC.CA> (24 Apr.1997)
- ⑦ Terry Kuny, <Terry.Kuny@xist.com> "Second IEEE Metadata Conference," 2 Dec.1996, <DIGLIB@INFOSERV.NLC-BNC.CA> (2 Dec.1996)
 - ⑧ A.Perkins, "Developing a Data Warehouse"
<http://www.ies.aust.com/~ieinfo/dw.htm>
 - ⑨ 陳雪華，「網路資源組織與Metadata之皓展」。《圖書館學刊》12期（民國86年12月），頁21。
 - ⑩ Marilyn Drewry, Helen Conover, Susan McCoy, and Sara J.Graves, "Metadata : Quality vs. Quantity," *Proceedings of Metadata Conference* " 97.
<http://computer.org/conferen/proceed/meta97/papers/hconover/mdrewry.html>
 - ⑪ David Levy, *Cataloging in the Digital Order, Digital Libraries '95* <http://csdl.tamu.edu/DL95/papers/levy/levy.html>
 - ⑫ Clifford Lynch, Avra Michelson, Cecilia Preston, and Craig A. Summerhill, *CNI White Paper on Networked Information Discovery and Retrieval*, Incomplete Draft <http://www.cni.org/projects/nidr/nidr.html>
 - ⑬ 同註⑩。

特別感謝

台大資訊工程所系統實驗室葉建華同學提供有關台大電子圖書館與博物館系統設計方面之資料，讓本文得以順利完成。

(收稿日期：民國八十七年三月二十日)