

《Metadata 專題》

中文 Metadata 的規劃與實作試驗探討

中央研究院 MetaData 工作小組第一、二階段報告

陳亞寧 江惠英 陳淑君 劉嘉慧

中央研究院 MetaData 工作小組的成立與任務

MetaData 工作小組於民國八十七年八月成立，目的在於針對院內進行中、或即將進行的數位典藏計畫，著手探討適合本院博物館、圖書館及檔案館等單位典藏資料的資訊描述格式與元素（format and element），以及檢索等相關事宜。依文獻及各研究計畫現況報導，國際間針對博物館、圖書館及檔案館所研訂的 Metadata 著錄格式雖有 CIMI（Computer Interchange of Museum Information）、CDWA（Categories for the Description of Works of Art）、EAD（Encoding Archival Description）、FGDC（Federal Geographic Data Committee）及 DC（Dublin Core）等二十餘種，然而這些著錄格式為求完整化、標準化及國際化等目標，目前仍屬持續研修階段，未臻完成。就現況研究需求與資料內涵而言，因本院館藏資料特殊，似乎沒有一種標準與著錄格式能完全滿足院內現有典藏資料特質的需求，加上 Metadata 議題的發展在國內外仍處於探索階段，工作小組採取對既有文獻及現有 Metadata 標準與格式進行研析，並掌握世界各項 Metadata 相關研究計畫發展現況，進而評估、分析，期能建構一套適用於本院多元化及臺灣本土化典藏特質的 Metadata 格式與元素。更甚一步，能兼容並蓄所有傳統媒體及由傳統媒體轉為數位典藏等兩部份的資訊內涵。

因此，工作小組主要任務有三：

- 制定符合本院進行的數位博物館專案計畫需求的 Metadata 格式。
- 以本院 Metadata 格式及實作經驗為基礎，研擬適合國內本土化中文資料特殊需求的 Metadata 格式與元素。
- 協同資訊及學科專家，共同設計網路化實際運作的作業系統（Metadata Testbed System）。

研究方法與範圍

工作小組係以印刷等不同傳統媒體及由傳統媒體轉為數位典藏為主要對象，準備規劃以下列方式、程序進行 Metadata 格式之研究與設計：

1.現有文獻研析及 Metadata 標準格式與元素分析

蒐集 Metadata 主題相關的文獻，由工作小組成員分別鑽研各種 Metadata 格式，並給予簡報。目前已進行八種 Metadata 相關範疇，包括：基本概念導覽、廣為使用的 Metadata (DC)、政府機構文獻的 GILS (Government Information Locator Service)、博物館與檔案館相關的 Metadata (CIMI, CDWA, EAD) 及整合各式 Metadata 的基本架構 (Warwick Framework, RDF (Resource Description Framework))。這些歷次的研讀心得將陸續於通訊中予以刊載，讓有興趣或關心 Metadata 應用於數位典藏資源的同道一起分享。

2.瞭解本院各單位資料庫系統建置情形

本院各所研究人員為了學術資料使用的方便性，早已建置符合本身需求的資料庫系統。無論是紙本、卡片式或是電腦型式，工作小組成員先行瞭解現有系統運作方式及其優缺點，作為設計 Metadata 格式及系統需求的參考。

3.實際訪談瞭解資料特性與使用者需求

實際與本院使用者及資料管理者進行晤談與溝通，以深入瞭解並確定雙方需求，同時經由資訊科技人員的參與，作為研擬 Metadata 的主要參考依據。目前本院在「國科會數位博物館專案計畫」中參與三項學科主題計畫，包括：臺灣原住民－平埔族群、不朽的殿堂－漢代的墓葬和文化、臺灣的本土植物和魚類。經由每週的定期會議，分別邀請計畫相關成員齊聚一堂，溝通、確認要求，並收集所有相關的資料屬性，讓 MetaData 小組成員更瞭解其所要表達的資料內涵和深廣程度。

4.制訂適合本院需求的 Metadata 資料表格、格式、原則與規範

本院為國內最高學術研究單位，在執行「國科會數位博物館專案計畫」之際，為能將豐富的學術資源呈現於網路上，以達到社會教育化的功能，因此對於既有的建檔資料與格式，必須參照國際標準加以規範制訂。

5.院內實作修正

工作小組以「臺灣原住民－平埔族群」計畫為首例，進行資料內涵的分析及屬性的確認，並模擬設計出 Metadata 輸入表單。經公開解說後，請該計畫相關學科之專家詳實確認、試填、實作，以及提出修正意見。

6.調查臺灣本土典藏資料特性及系統建置實際情形

7.參加國內外 Metadata 相關主題之研討會

為使 Metadata 的規劃作業更臻完善，工作小組成員目前已出席參與兩項國際 Metadata 相關會議及研討會：「1999 PNC (Pacific Neighborhood Consortium) Joint Meeting」、「數位博物館資訊組織研討會 (Workshop On Dublin Core for Museums)」。

8.完成系統設計與運作雛型

9.舉辦公聽徵求意見，以完成系統設計與實作

Metadata 設計的目的導向

工作小組針對本院數位典藏計畫的內涵進行分析，對 Metadata 設計的主要目的有如下八項：

1.資料架構與模式 (structure & model)：設計一個共通性組織結構，以容納不同資料類型與學科領域的 Metadata。

2.資料輸入與描述整理 (input & descriptive organization)：為典藏品資料建立一套詮釋性的記錄。

3.檢索與索引 (retrieval & indexing)：讓使用者很有效率地進行查詢這些記錄。

4.展現與辨識 (representation & identification)：從查得的記錄中，使用者可以清楚地獲得所需的訊息及制訂呈現方式。

5.串聯與互動關係 (linkage and interactive relationship management)：建立不同文獻間的串聯架構、方向 (雙向與多向)、模式與管理等。

6.取用與認證 (access & authentication)：作為系統安全控制的機制功能之一，以區分不同身份的使用者，包含智財權 (intellectual property rights) 的管理與控制。

7.交換與儲存 (interchange, mapping & exchange and storage)：這些詮釋性記錄可以因各種不同需求 (包括書目與全文兩部份) 而進行交換及儲存。

8.整合 XML (eXtensible Markup Language)、RDF 與 Z39·50 不同協定的應用：除了致力於 Metadata 的制訂，因應文獻結構的制訂、交換、檢索與展現的需求，另結合 XML、RDF 與 Z39·50 等協定的應用，以發揮 Metadata 的功能。

Metadata 設計原則

基於本院典藏資料及文物的調查和研析，為能建構一套兼容台灣本土典藏的特質，又能兼顧本院學術研究需求的功能，最後並能符合網路資源的國際化，工作小組乃訂定以下八項原則，作為規劃本院 Metadata 之依據：

1.符合國際標準要求，包括 Metadata 格式、網路協定、檢索 (如 Z39·50) 等。

2.依據本土需求採用既有標準進行修改，不再另行重新研發。

3.容易使用，包括資料的建立、展現、檢索、詮釋與串聯關係的互動。

4.具備多語文能力。

5.因應不同學科領域需求採用不同的 Metadata 格式，而非試圖設計一套單一、一致性的 Metadata 格式，以容納不同的 Metadata 格式於一身。

6.採用不同 Metadata 格式，同時利用 XML 建置 Metadata 基礎骨幹與結構來容納、整合不同的 Metadata 格式，包含現有圖書館自動化系統的結合，以利一次查詢所有館藏資源。

7.設計系統管理機制功能以因應不同的需求，包括取用限制、智財權的保護。

8.具有相當的延展性，包括不同資料格式間的轉換、儲存與展現，以及資料欄位的制訂與修正。

Metadata Record Sample

物件種類	古文契書
契約類別	契尾
原件編號	2184-2
語文別	中文
契字名稱	契尾布字貳百貳號
立契人	魏慶雲
立契時間	咸豐元年
立契地點	舊地名 淡水廳九芎林庄
今地名	新竹縣芎林鄉
類別/主題	地契(契尾)
實體描述	1 張 印刷樣式 章 5 枚 右旗縫章 2 枚 右款縫 魏慶雲收執 文字破損 其他: ET2184-1 之契尾;金廣福文書 029
系統管理	建檔者/日期 sophy/19990222 審核者/日期 arthur/19990222

最新審核日期 19990222

最新更新者/日期

sophy/19900222

Metadata 的設計架構

工作小組期望設計 Metadata 的原則，係本著使用容易、詮釋深入為目標，因此在設計架構上希望達到如下的特色：

- 格式的整體架構

可由核心集與附屬集組合而成，讓資料保有者可視資料特性，以決定所要著錄的層次。例如，最小的核心集可設置 10 個資料元素，附屬集可更詳盡地再設置細項，以更進一步來詮釋項目的內涵。

- 資料著錄元素

工作小組制訂的記錄元素除了應用於資料建檔外，更有不同層次的應用與詮釋（詳見文後” Metadata 記錄架構及輸入表單設計”之 Metadata 記錄元素與展現說明）。

- 文獻關係的呈現（context relationship）

所謂「文獻」包含紙本式的一般文件及各式影像（如圖片、動畫、聲音等），而「關係」在此則包括如下幾種情況：母子作品的關係（如套書）、作品前後的關係（如上下集作品）、作品斥補的關係（如可作為補充的作品，或不同觀點的相關作品及版本等）。若這些文獻的關係得以盡可能的詮釋出來，則文獻間的呈現將更具互動性、貫穿性與完整性。

Metadata 記錄架構及輸入表單（template）設計

基於前述預期設計達成的特色，有關 Metadata 記錄架構與輸入表單方面，分述如下：

- Metadata 記錄架構

基本上分爲主要元素（core element）及從屬元素（individual element）兩種層次。所謂主要元素乃是不同 Metadata 格式間所共通、共用的資料欄位；而從屬元素則是各式 Metadata 格式所獨有的資料欄位。但是一個完整的 Metadata 格式是由主要及從屬元素共同組成，如圖一所示。

• Metadata 記錄元素

在工作小組著手完成制定的記錄元素，主要是作爲系統資料庫內部架構建置與運作的主要依據。爲了達成高度的親和性、彈性與廣度，就資料輸入表單而言，工作小組直接採用資料內涵專家熟悉的術語作爲輸入顯示欄位名稱（label），例如古文契書，而不採取記錄元素爲欄位名稱。就資料轉換與交換方面而言，亦以此作爲互轉的依據。此外，爲求取資料能放置在 Web 介面上，也以記錄元素爲基石轉換成 HTML 檔。上述這些不同層次的應用與詮釋，請參見表一及表二實證說明。

研究發現與建議

發現 1. Metadata 評選原則與要點

目前全球已有各式 Metadata 格式，但是本院在設計適合本院需求的 Metadata 時，仍不宜貿然採取「從無到有」的自行研發方式，改以採用「現有標準」爲首要原則再進行修改。Metadata 工作小組成員係以瞭解各種 Metadata 格式著手爲切入點，再依據對院內現有數位典藏計畫資料內涵進行研析，提出下列要點（criteria）作爲評選 Metadata 的準則依據。評選要點如下：

1)使用環境

- 使用者族群特性：對象、學科
- 資料類型：printing, audio, video, image, photo 等
- 國際標準符合程度與相容性：網際網路、圖形介面（GUI）、資料格式
- 多語文
- 應用容易度：設計者、管理者、使用者

2)格式

- 結構

- 資料交換與互轉
- 文獻互動關係之建立與管理
- 內容符合比率

3)系統功能要素

- 建立 (input)
- 索引與檢索 (index)
- 展現 (representation)
- 儲存 (storage)
- 交換與互轉 (exchange & interchange)
- 其他 (others)

發現 2. Metadata 記錄架構初步證實可行

工作小組所設計的 Metadata 架構經以臺灣原住民－平埔族群計畫中的相關資料實作後，初步發現可行性相當高，但由於目前院內尚有其他兩項主題計畫（臺灣的本土植物和魚類、漢代的墓葬和文化），可能涵蓋的資料類型與特質會有不同。因此，更詳盡深入地瞭解這兩類型的資料特性以及臺灣原住民－平埔族群計畫的其他資料特性（如，聚落資料），將可更詳實確認架構的可行性。

發現 3. Metadata 設計及執行之效率問題

工作小組在設計輸入系統介面 (input editor) 時，係以呈現出不同主題的特有語彙為導向，讓學科專家在熟悉的環境中建立 Metadata，因此不同資料類型可能需有不同的輸入表單與欄位名稱。例如：專書、單篇文章、古文契書、地圖資料等。同時為顧及每一種資料類型的獨特性而為之量身訂做，在建置的效率上是一大考驗，因此適當地整合會是一個重要的工作要點。

建議：

下階段的工作重點，工作小組將建議除持續與學科專家 (content experts) 晤談、溝通外，Metadata 權威控制的建立當屬首要之急，期使以 Metadata 的呈現更具質感，資料能夠更精確地被記載、詮釋與串聯。例如人名權威檔、地名權威檔等；另一方面，如何更有效率地進行 Metadata 的建置作業，亦是需要改善之處。最後，以系統建立的角度而言，資料間之互動模式也是極待發展與建立的重要工

作。