

《Metadata 專題》

Metadata 設計分析與資料模式、物件導向的應用互動

中央研究院 MetaData 工作小組第三階段報告

陳亞寧 陳淑君 江惠英

前言

在制訂 Metadata 格式 (format) 與各項元素 (elements) 時，除了著重在研析、應用現有的各式 Metadata，或自行創立新式 Metadata 的考量外，更重要的是在確定學者專家的需求，以及資料類型與特質後，如何依循一套有效、正確的研擬方式來設計、確定與因應 Metadata 格式、元素，是一項更為重要的原則與方向。本文試以本院數位典藏計畫中的平埔子計畫為例，探討如何應用物件導向 (Object Orientation, OO) 與資料模式 (data model) 等方法進行 Metadata 的研發、設計與應用。此外，為能適切表現所謂動態資訊與知識 (dynamic information and knowledge)，針對各項資源的互動關係亦進行探討。本文主要研究重點有：Metadata 範圍、Metadata 結構與應用、Metadata 應用層面、資源互動關性等四方面。

Metadata 範圍

目前各式 Metadata 的設計制訂，所涵蓋範圍大部份以既有傳統型式轉為數位型式的數位資訊載體為主，當然也有以現有的傳統媒體（如印刷、標本等）為應用範圍【Miller, 1996】，以及本身即為數位型式的各類資訊等三大類型。以 Dublin Core 為例，在一九九五年三月第一屆 Dublin Core 會議中，Dublin Core 涵括對象是以文獻 (document-like-objects) 為主要範圍【Weibel, Godby, and Miller, 1995；Weibel Iannella, and Cathro, 1997】。至一九九六年九月第三屆會議中，又將影像資料 (image-like-objects, non-document-like-objects) 納入 Dublin Core 範圍中【Weibel and Miller, 1997】。如果經由文獻探討，可得知現今各式 Metadata 最初目的在於解決網路資源探索 (network resources discovery) 的問題。但是隨著資訊科技日新月異、資訊載體多元化及使用者需求多樣化等前提下，Metadata 的設計導向已逐漸擴展至

不同領域、層面。總而言之，Metadata 包含範圍是以現有存在的各類資源 (resources) 為對象，而每一個或是每一種資源皆是一種物件 (object)。從圖書館、檔案館、博物館與美術館觀點而言，現有的圖書、期刊、手稿、標本、畫像、雕塑品等皆是。依本院 MetaData 工作小組 (以下簡稱「本小組」) 針對本院現有的數位典藏計畫 (平埔、墓葬、植物與魚類) 資料內容而言，所謂的物件群組 (class) 可分為物 (object)、人 (people)、時 (time)、地 (space) 及事件/主題 (event) 等五大類型。亦即，從文獻與影像等既有範疇，轉換、擴大為以物件為範圍 (請參見圖一)。若以物件導向設計觀點而言，每種群組下，由上而下又可區分為大物件 (object) 及小物件 (instance) 等不同層次。例如，在「物」這個群組下，可以區分為書、期刊、單篇文章、銅鏡、璽印等不同大物件；而每個大物件下，則是每個小物件 (如每一本書，每一種期刊、每一面銅鏡、每一顆璽印) 共同組合而成。除此之外，每一個小物件相同層次間即有不同的互動關係，並向上擴展至大物件及群組層次，形成另一種虛擬式互動關係。換言之，在物、人、時、地相互交錯下，除了本身即具備互動關係外，更重要的是，可能在物、人、時、地等其中的二、三、四個主軸共同牽引、互動下，引發事件/主題的產生。例如，某一特定平埔族聚落的產生、遷移、改變、融合與消失等現象皆屬之 (請參見圖二、三)。

Metadata 結構與應用

以下將分 Metadata 記錄與元素、記錄與各式 Metadata 的相互關係、Metadata 元素的設計與實作等四方面說明如后。

1. Metadata 記錄與元素

本小組在設計適應於本院數位典藏計畫內涵需求時，並非自行研發一個新的 Metadata 格式，而是利用現有的各式 Metadata 為基本原則。因而本小組為了擷取各式 Metadata 元素，就必須針對記錄結構方面加以研析、設計。對此，本小組將之劃分為核心 (core) 與個別 (individual) 兩部份 (請參見圖四)。前者乃是每筆 Metadata 記錄共通使用的元素，並不因學科領域或是資料類型的不同而有所區別；後者則是為了因應不同學科內涵與資料特性等需求所獨有、特殊的元素。目前核心元素是取自於 Dublin Core，而個別元素則源自於各式 Metadata，或是無法取自各式 Metadata 時，本小組則自行設立。例如系統管理、智慧財產權控管等方面的元素。就

整個記錄結構而言，兩者共同組成一筆 Metadata 記錄，而且兩間彼互通，一旦核心元素不再具備共通性時，則元素自動轉換為個別性質；反之亦然。

2.記錄與各式 Metadata 的相互關係

基本上，本小組並不試圖設計一套共通性的 Metadata 格式，以整合各類 Metadata。換言之，即是全球共通的 Metadata (universal metadata format)，或是超級的 Metadata (super-metadata) 格式

【Chilvers and Feather, 1998】。定位若非如此，本小組制定的 Metadata 何以能與現行各式 Metadata 相通呢？主要原因有二：一為，從記錄結構與元素而言，本院並不使用所有 Metadata 格式及其元素，而是擷取所需的元素，並經由一種有彈性的基礎架構以組成 Metadata 記錄結構，以達到不同的應用與融合；二則，現行各式 Metadata 雖然結構、內容、元素不同，但是經由架設的記錄結構為基準，內外不同 Metadata 格式及元素仍可由記錄結構達成互通。即透過不同的轉換表及程式 (mapping table and program) 達成無礙地互轉，以形成一種透通性的內外轉換與交換。

3.Metadata 元素的設計：以物件導向為主

就整個記錄結構而言，雖然分為核心與個別兩部份，但是各元素是同等視之。即每個單元皆視為各別獨立的物件，以達成物件導向與記錄結構化，並不因元素不同而作不同的處理。以“作者”為例，在設計時分為作者、著作方式及內容 (value) 等三部份，其中著作方式若已有現行標準 (如 USMARC Relators) 則直接採用 (詳見圖五)。此外，本小組在資訊技術人員的協助下，亦採取資料模式方法設計、規劃各項元素及整體性資料庫，期使資料元素更具結構化，以及建立彼此間的互動。以平埔計畫為例，本小組首先與內容專家進行訪談，接著再充份瞭解各項需求與分析出各項元素後，再由資訊技術人員經過歸類、分組各項元素屬性，次而進行彼此間的上下隸屬、互存與互斥，以及互動關係特質等正規化 (normalization) 程序後，制訂出各項元素及資料庫基本架構 (請詳見圖六至七)。

4.實作說明：以平埔計畫內涵為例

以平埔族計畫為例，從所屬資料內涵而言，原本從研究平埔學者

提供的資料類型可分為研究書目、全文、地圖、契約文書、影像、訪問、錄影與音樂等八種型態，但經過內容屬性分析與研判後，其實可分為圖書等廿二種類型；然而就系統內部而言，根本上只有一種，唯可依使用者需求，提供設計成學者專家日常研究工作中所使用、且最為熟悉的術語，組成一項資料輸入表格（**template**）及欄位名稱（**label**）（請參見圖八標示），可謂達到「本尊唯一，分身無數」的現象。

Metadata 應用層面

以 Metadata 整體應用而言，本小組並不將 Metadata 僅僅視為描述資料外表而已，或是依循圖書館行之多年分編理論與作業觀點，只將 Metadata 作為外形描述，以及利用主題標目與分類號等系統標引內容屬性而已。本小組從輸入、轉換與交換、顯示、索引、串聯、儲存，以及不同 Metadata 間整合等角度整體觀之，而且資料型態也從原有的書目資訊，進而擴大至目次、摘要與全文等三方面。有關利用 Metadata 詮釋各種學科領域的資料與知識內涵方面，除了針對原本資料外表型式的正確描述與主題屬性的適切標引外，當然對於人、時、地、物及事件/主題也進行整合、建置與架構，以求同時詮釋靜態（**static**）與動態（**dynamic**）的資訊，進一步適當地詮釋相同與不同的群組、大物件及小物件間互動關係的建置、標引、說明與詮釋（詳見後文“資源互動關聯性”說明）。

至於 Metadata 的輸入、索引、顯示、儲存、交換、串聯等應用，本小組是應用 RDF（**Resources Description Framework**）與 XML（**eXtensible Markup Language**）等協定與觀念來直接或間接達成整合（請參見圖九）。在輸入與儲存方面，終端使用者雖然是經由圖形介面輸入與維護，同時經由本小組的 Metadata 元素名稱作為欄位名稱，但是居於介面與 Metadata 元素間則是 RDF 與 XML，只是終端使用者無法察覺而已，但儲存格式則為實體的 RDF 及 XML。至於不同 Metadata 間的轉換與系統間的交換，則是將 RDF 與 XML 視為轉換介面，經由兩者達成內外間的互轉與交換，同時亦作為容納各式 Metadata 的基礎（**framework**）或封包（**container**）。有關索引與檢索方面，全部皆以本小組制訂的 Metadata 元素為主要依據，經由學者專家決定後，由 RDF 與 XML 整合不同的 Metadata 格式與元素，同時加以包裝後，再經 Web 介面提供終端使用者查詢使用。至於串聯方面，最主要是透過 XML 功能直接完成，其餘詳見後文“資源互動關聯性”說明。

資源互動關聯性

以平埔實作中，採取物件為導向的依據下，已區分為人、物、時、地及事件/主題五大主軸。在不同主軸間對交錯激盪下，其實不同資源間存有某種的互動、關聯性。就本體觀點而言，各項資源內涵、承載的資訊與知識本為靜態式虛擬本體，即使經由 Metadata 描述與整理後，仍為靜態不動的，但是經由互動關聯性的標示與建立，原本靜態資訊與知識得以模仿人類既存的認知模式進行呈現，以達成所謂的動態資訊與知識。換言之，經由此種模仿人的智慧組織型態與知識體系網絡等模式，達成知識本體的建構與展現，本小組已歸納出九大組別，茲分別闡述如下：

1. 母子關係 (part and whole relationship)

所謂的母子關係，就實質上或邏輯上而言，即某一資源隸屬於另一資源的一部份（如圖十所示）。例如，套書、叢書、期刊與單篇文章、論文集與單篇論文、研討會論文集與單篇研討論文等。

2. 重生關係 (creative relationship)

乃是某一資源與另一資源之間的各种衍生或再創關係。通常是依據某一資源的內涵精神，再以另一種方式（表演、複製、翻譯、改編、詮釋，或改變形式的派生）來重新表達或詮釋，因此已儼然形成一種全新的資源，即稱為重生關係（如圖十一所示）。應用範圍包括翻譯作品（translations）、改寫或改編作品（adaptations or arrangements）、樣式改變的作品（changes of genre，如小說改成劇本）、模仿作品（根據原作品精神風格予以自由譯述、意譯、模仿或諷刺模仿的作品）等。

3. 版本關係 (version relationship)

版本關係乃是某一資源與另一資源隸屬同一作者，並有版本方面的歷史淵源等前後演變關係。亦即，透過修訂、擴大或節縮等方式變化原作品的產物，例如修訂版、原稿（draft）等，如圖十二所示。

4. 形式關係 (format transformation relationship)

某一資源藉由複製 (reproduce) 或重新組成 (reorganize) 等技術，在形式上被改變成另一資源。此處資源間的改變係為一種外觀或型式上的再改變 (transformation) 與呈現 (representation)，而非內涵精神方面的重新詮釋，此即稱為形式關係。範圍包括任何內涵 (intellectual content) 及其著作方式 (authorship) 相同，但呈現不同型態外觀的作品之關係，如圖十三所示。例如，複本/刷次本 (copy/impressions)、版次/重刊 (issues/reissues)、傳真本/再印本 (fascimiles/reprints)、影印本 (photocopies)、微縮複製品 (microform)、其他複製品 (包括 format change、mirror copy 等)。

5. 參照關係 (reference relationship)

所謂參照關係，乃是某一資源內容中引用 (citation)、感謝 (acknowledgement)、評論 (review)、參考 (reference) 或提及另一資源，而形成資源間的參照關係，如圖十四所示。例如，作品的參考文獻、書及其書評、作品的註解與解釋版本、文學評論等。

6. 依附/共生關係 (dependent relationship)

某一資源必須倚賴另一資源為基礎後，方能運轉、傳遞，或是另一資源完整內容的一部份。通常這兩個資源間有主、次要之分，功能上彼此可以相互補充，以突顯主要資源所要呈現或表達的主要意念，如圖十五所示。型態可分為補正主要資源的內涵 (如修正、增補)、補遺 (如附錄、使用指引)、圖解說明主要資源 (如地圖集、插畫)，以及提高使用主要資源的資料 (如索引) 等四種不同的類型。

7. 連續關係 (sequential relationship)

連續關係係指在時間的前後時程上，某一資源與另一資源間的關係是具備連續與繼承的特質，範圍涵蓋了合併、分裂、易名、衍生等不同層次，如圖十六所示。例如，連續性作品 (serial) 改變刊名後的新舊刊名作品間關係、作品的後續作品 (sequels，如續集)、叢刊作品 (series) 各部分作品之間的關係。

8. 隸屬關係 (hierarchical relationship)

某一資源與另一資源關係呈現著隸屬關係。換言之，兩者間存在著某種上下附屬層次的關聯性質，此種關係即是所謂的隸屬關係，如圖十七所示。例如，教育與教育行政的關係。

9. 共通關係 (shared characteristic relationship)

共通關係乃是指各資源間看似並無任何的實質關係，但是卻存有一種透通式的共同性質，將這些毫無關係的各項資源予以串聯、集合在一起，以引發、產生另一種新的關聯性，例如前述的聚落（請參見圖四）。

結語

經由本小組以平埔計畫內容為範例，以及採取物件導向與資料模式等方式基礎下，目前刻正進行各項 Metadata 的設計、轉換、整合等不同層面的應用，已初步證實本小組在不另創立一種新的 Metadata 格式與元素，以及應用既有各式 Metadata 格式與元素等兩項前提下，結果是可行的。同時在配合資源互動關聯性關係的建構下，原有靜態的資訊與知識亦可成功地轉換為動態型式。但是有關 Metadata 與 XML 及 RDF 方面的結合應用，仍待下一階段的實作與求證。（備註：本文承蒙本院資訊所王友橋及鍾子帆兩位先生協助提供有關文中圖六與圖七等資料，方能順利完成，在此一併致謝。）

參考書目：

1. Chilvers, A. & Feather, J. (1998). The management of digital data : a metadata approach. *The Electronic Library* 16(6), 365-371.

2. Paul, M. (1996). Metadata for the masses. *Ariadne* (5). [Access Date : 20 April, 1999]. Available : <http://www.ariadne.ac.uk/issue5/metadata-massess/>

3. Weibel, S., Godby, J., & Miller, E. (1995). OCLC/NCSA metadata workshop report. [Access Date : 31 May, 1995].

Available : http://www.oclc.org:5046/conferences/metada/dublin_core_report.html

4. Weibel, S., Iannella, R., & Cathro, W. (1997). The 4th dublin core metadata workshop. [Access Date : 19 June, 1997]. Available : <http://www.dlib.org/dlib/june97/metadata/06weibel.html>

5. Weibel, S., & Miller, E. (1997). Image description on the Internet : a summary of the CNI/OCLC image metadata workshop. [Access Date : 20 August, 1998]. Available : <http://www.dlib.org/dlib/january97/oclc/01weibel.html>