

UML 應用於數位典藏後設資料之設計

蔡順慈

sttsai@faculty.pccu.edu.tw

中國文化大學資訊傳播研究所

林昱伍

fast.lin@msa.hinet.net

摘要

有鑑於傳統文化保存的重要，世界各國皆先後推動數位典藏計畫。數位典藏除了必須有效呈現文物的數位化內容。還需要有完整額要的典藏品描述。目前各典藏單位致力於後設資料(Metadata)的建立，詮釋文物的重要性與保存知識。

數位典藏文物描述使用的國際標準規格元素欄位，做為後設資料綱要的設計。在儲存方面以資料庫欄位來設計，利用關聯式資料庫(Relational Data Base, RDB)做為儲存欄位設計，對映輸出為XML的檔案格式以供交換查詢之用，為目前主要的設計方式。

本文提出以UML做為後設資料綱要的設計，提供對於RDB與XML之不同資料呈現結構的轉換，增進在設計與修改上更有利於溝通與維護，期望以此架構做為後設資料系統分析與軟體開發之參考。

關鍵字：後設資料、數位典藏、統一塑模語言、可延伸式標記語言

Abstract

As the traditional cultures preserve, each country expand digital archive program successively. Digital Archives present collections really. Besides, it needs to describe completely. Each organization creates Metadata more effort to provide knowledge preserve. The collections of Digital Archives use international standard metadata element to describe for metadata schema design which uses database field and RDB (Relational Data Base) in store view. It correspondents with XML to exchange. This is main method for metadata establishing.

In this paper, use as UML to establish Metadata schema to provide for RDB structure transforming XML structure. It's easy to understand and maintain in maintain and modify. It expects to metadata transform on software design.

Keywords: metadata, digital archive, UML, XML

1. 前言

資訊科技的快速進步，對於人類重要智慧結晶的文物保存必需要隨時代精進而更利於推展。對於檔案呈現的品質與長期的儲存的格式以及快速檢索的技術，也為目前各國所極力發展數位典藏的重點[1]，數位典藏(digital archiving)，意指長期的電子資源儲存、保存及取用，將原生的數位資料或是原始資料數位化後的資料，以數位資訊的形式更易於傳播、複製及再利用[9]。

全球重視傳統文化的國家，多已致力於數位典藏計畫的發展，先後推動數位典藏，提供許多相關技術和針對各國最具代表性的典藏品以數位化的方式長期保存，其中包括了人文、藝術、歷史方面等，建立數位化的圖文資料庫，將珍貴的人類知識與智慧結晶加以實行數位化工作，方便更有效的傳播應用[10]。

2. 數位典藏與後設資料(Metadata)

數位典藏除了以數位化方式呈現文物的擬真原貌，還需要有完整的典藏品描述文件。而目前文物特色的闡述是以後設資料(Metadata)建立，對於後設資料的描述格式與數位化的影像品質為建立數位典藏重要兩大關鍵。

2.1 後設資料意義

目前各計畫對於數位典藏的執行，都積極於後設資料的建立，Metadata的基本定義出自OCLC與NCSA所主辦的「Metadata Workshop」研討會的發表，國內翻譯名詞尚未統一，在資訊組織界最普遍的解釋是“data about data”，意思為描述資料的資料[7]。

如今後設資料的理念不再只是描述文物特性，後設資料主要任務可提供解說特定領域的相關知識，而未來後設資料的發展會實現各領域專業知識的關聯，並以媒體整合為基礎的知識空間[4]。

2.2 後設資料設計

數位典藏的後設資料在設計上必需考量文物本身的特性來選用適合描述的後設資料格式，分散式資訊的整合檢索，雖不是後設資料的直接內涵，但在網路環境中也是一個重要的課題，而且相關標準的設計和後設資料息息相關[7]。

針對特定領域資訊的共同需求與著錄標準，各領域已訂定不同的後設資料標準格式，例如：典型代表圖書館社群：MARC，博物館社群：CDWA，檔案館社群：EAD，視覺資源社群：VRA Core Categories，植物標本館社群：HISPID，然而 mda(The Museum Documentation Association)所發展針對博物館的後設資料格式，SPECTRUM (The UK Museum Documentation Standard)也是廣為使用的後設資料，包括歐洲各國、大英博物館、CIMI 博物館數位典藏計畫等均有參考。

另外網際網路普遍採用的 Dublin Core 後設資料規格，以較適用欄位格式，供交換查詢之用，在輸出時最多採用 Dublin Core 的 15 個欄位，對映後以 W3C 建議的 XML (Extensible Markup Language) 檔案輸出，捨棄 SGML 過於龐大複雜以及不易普及化的缺點。

決定了資料應描述屬性後，接下來的問題就是建立屬性的結構。屬性的結構在呈現屬性之間的關係，使得屬性的描述清晰易懂，以 XML 檔案格式來呈現資料，然而 DTD (Data Type Definition)所規範的格式即是這個層次的問題，也決定了後設資料綱要(Metadata Schema)，DTD 的作用在於定義和規範特定 XML 文件的內容架構，它通常是一個含有某一種特定格式正式定義的檔案文件。

3. 後設資料結構呈現問題

針對於後設資料的系統設計，主要是依後設資料功能需求表單的項目來建立，而後設資料綱要為關於描述典藏品所使用的國際標準規格的元素欄位，在儲存方面，以資料庫欄位來設計，一般是利用關聯式資料庫(Relational Data Base, RDB)來作資料儲存的處理，而後設資料綱要為和 DC 對映輸出為 XML 的檔案格式以供交換查詢之用。

由下圖 1 主要說明以 RDB 與 XML 設計上的差異性，(a)RDB 是由表格的記錄所組成，而(b)XML 的資料卻是以樹狀結構來表達[5]，所以在管理和對映資料上相當複雜。

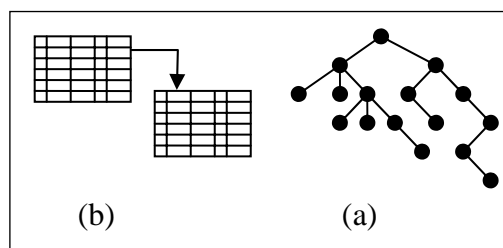


圖 1 XML 與 RDB 資料結構表示

兩者之間的結構有著極大的差異，這造成利用 RDB 來管理 XML 文件一些新的挑戰，為了要能方便處理這些 XML 資料，幾乎所有的資料庫廠商都已提出了簡單的解決方法。

4. 應用 UML 呈現 Metadata 設計

統一塑模語言 (Unified Modeling Language, UML) 乃基於物件導向觀念發展而來的塑模語言(Modeling Language)，它除了具備物件導向特性之外，統一塑模語言更從使用者、系統開發者等多重觀點來架構出一套完整的分析模式語言。

數位典藏整體工作需大量的人力、技術等各項不同的工作流程與資訊科技，需要使用深入簡出的方法，解釋整體的執行步驟與架構，期望在整體工作的發展過程中，不同領域的專業人員都可用此標準語言來進行溝通。

要交換 XML 文件，就必須使用綱要語言來制定 XML 語彙，在 OMG (Object Management Group)的 XMI (XML Metadata Interchange) 規格書有針對 DTD 的轉換對映說明，包括轉換為 W3C 的 XML Schema，因而由此衍生出 XML 綱要設計之問題。UML 資料模型與 XML 綱要之間轉換的議題，發展出一個可進行雙向轉換的模型。

以 UML 模型將分別探討 XML DTD，分別從語法及語意轉換之層面切入進行一連串結構轉換規則之延伸及建立，在 XML 綱要結構反向轉換為 UML 資料模型結構之模型，分別由 XML 綱要結構建立轉換之規則。

建立此轉換模型之目的在於使 UML 資料模型結構及 XML 綱要結構間之轉換方法更為明確、簡單及容易使用。以博物館提供各種典藏品的後設資料，表 1 以 CDWA 來描述，而只將典藏品名資料對映 Dublin Core 做為交換查詢之用，以 XML 檔案格式輸出，而博物館的以品名與作品內容編目並提供線上圖片瀏覽，所以開始分析建立博物館的後設資料，分析後設資料需求表單的格式，以樹狀表格結構表示：

表 1 後設資料需求表單

CDWA		DC
品名 (Object Title)		Title
作品內容 (Subject)	主題意涵 Type	Description
	表現題材 Topic	Description
URN		Identifier

本文以 Rational Rose 軟體做為 UML 的設計工具，依後設資料需求表單，使用類別圖分析，在轉換格式將 Stereotype 改為 <<Table>> 與 <<DTDElement>>，因為考慮到後來的切割，所以可能會切割的元素為一個類別圖，而線上圖片是以透過網路連結的方式來提供典藏品圖片的查詢，所以繼承的方式來呈現。

而對映方式是依類別圖本身的結構，直接轉換為 XML 的元素與屬性架構，圖 2 所產生的 DTD 為產生文件結構的基本原則，而對於較為嚴謹的 XML Schema 的轉換亦是相同方式[6]。

以 UML 與 XML DTD 對映的方式，類別圖本身就記錄了屬性與元素，在轉換為 DTD 時，所表示的元素以 <ELEMENT> 標籤直接定義所包含的元素，並由 ATTLIST 標籤來定義所代表的元素屬性，以 XML Schema 來表示元素屬性方式亦也相同，只差別標籤的定義不同。

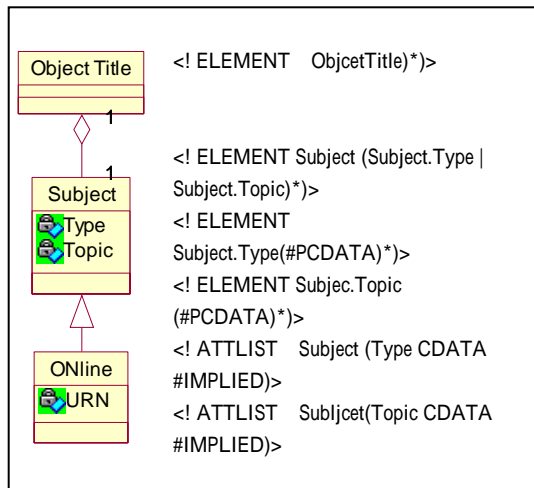


圖 2 UML 與 DTD 之轉換表示

而此架構對於 XML 文件的本身結構上，亦可呈現，此時各類別圖的關係對於文件的產生亦為非常的重要，UML 類別圖的結構包含了屬性、關聯的終點參照、與組合，在 XML 文件上的顯示方式，以關聯的終點看成是包含在類別裡。UML 類別圖主要說明了關係與屬性，這對於關連式資料庫的設計，提供了主要的關係說明，有利於資料庫的綱要的設計，許多工具皆有提供這兩者之間的轉換。

類別圖關於資料庫設計主要是將聚集

(Aggregation)、組成 (Composition) 和關聯 (Association) 關係以一對一、一對多、或多對多的方式。繼承關係在資料表設計上有許多不同的方式，本文以 Rational 公司定義繼承關係就是 UML 的一般化 (Generalization) 關係，在對映為資料表格時，是將父類別與各個子類別分別對映成資料表格 (實際數量為子類別的個數加上一個父類別)，父類別的主鍵將成為子類別資料表格的主鍵(primary key)或外來鍵(foreign key)，如圖3所示，而在要呈現的資料表必須自主/外來鍵，無法做統一性的轉換。

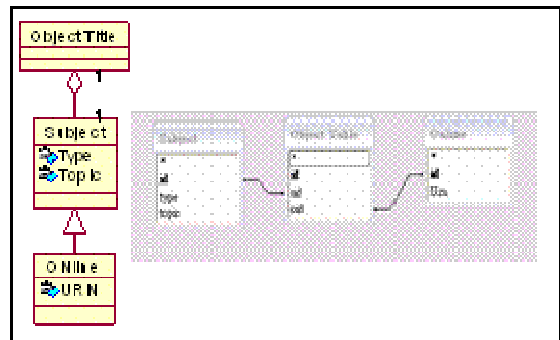


圖 3 UML 與資料庫綱要之對映

而此架構對於 XML 文件的本身結構上，亦可呈現，此時各類別圖的關係對於文件的產生亦為非常的重要，UML 類別圖的結構包含了屬性、關聯的終點參照、與組合，在 XML 文件上的顯示方式，以關聯的終點看成是包含在類別裡，圖 4 以一個油畫典藏品，品名為“山水畫”，用 XML 元素表示：

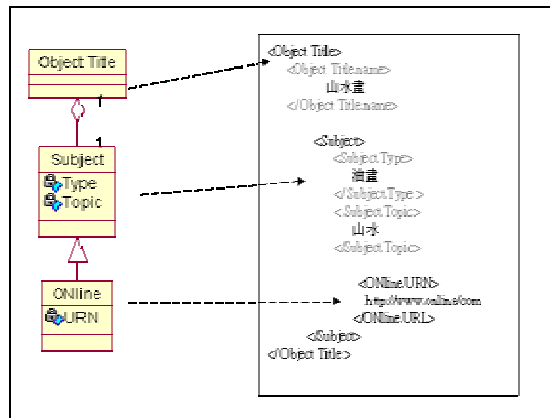


圖 4 UML 對映到 XML 元素

以 Dublin Core 的對映輸出的格式，由圖 5 所示將對映的資料欄位直接指向 XML 的元素表示，箭頭所指的 XML 元素，會包括指向的類別。

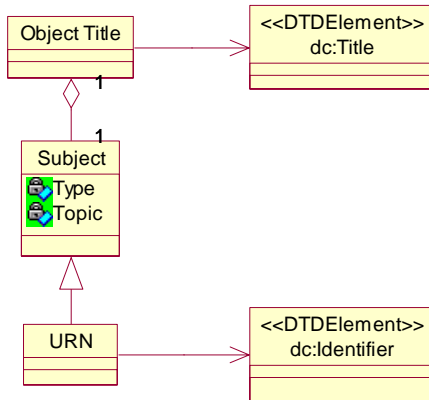


圖 5 Dublin Core 對映輸出

5. 結論與討論

5.1 UML 的結構表示與轉換

以 UML 設計後設資料轉換的各不同的格式，進而產生了交換的格式與資料表，大幅減少了開發的時間，並且在維護與修改上更有利於軟體開發人員溝通與維護，但在轉換設計上還是有不足之處，因 XML 一個元素以一個類別(Class)來表示，以數位典藏所用的後設資料規格，所定義的欄位皆為多個，在設計上會使整個類別圖非常的龐大，而資料庫也有必須切割成多個資料表，在關聯時的主鍵與外來鍵的設計上尚有缺失。

後設資料針對語意(Semantic)的表示結構，以 RDF(Resource Description Framework)是以圖形的結構表示後設資料，以 UML 可用狀態圖來表示，如果能和類別圖做統一的格式轉換，將更有效的以 UML 將這些主流的描述語言做一整合的表示，雖然設計上必須了解 UML 的表示方式，但可以依此架構做為軟體開發的轉換之參考。

5.2 原生型資料庫的應用

資料的關聯一旦要發生異動，則牽一髮而動全身，對應程式的修改是相當費時費力。有相關研究提出可利用原生型資料庫將整份文件儲存起來或是拆解成數個部分儲存起來，XML 的文件以原生型 XML 資料庫儲存，或許會是最自然和最佳的搭配，例如：X-Hive/DB、Xindice 皆是設計原生型資料庫的應用軟體。

誌謝

感謝行政院國科會及數位典藏國家型科技計畫補助本研究

(編號：NSC 93-2422-H-034-001)

參考文獻

- [1]. Ben Shneiderman, Meeting Human Needs with New Digital Imaging Technologies, *IEEE MULTIMEDIA*, Vol. 9, No. 4;2002, pp.8-14.
- [2]. Consultative Committee for Space Data Systems (CCSDS), Reference Model for an Open Archival Information System (OAIS). *Blue Book*. Issue 1. January 2002.
- [3]. David Bodoff, Mordechai Ben-Menachem, Patrick C.K. Hung, Web Metadata Standards: Observations and Prescriptions, *IEEE SOFTWARE*, Vol. 22, No. 1; JANUARY/FEBRUARY 2005, pp. 78-85
- [4]. Frank Nack, The Future in Digital Media Computing is Meta, *IEEE MULTIMEDIA*, Vol. 11, No. 2; APRIL-JUNE 2004, pp. 10-13
- [5]. S. Banerjee, V. Krishnamurthy, M. Krishnaprasad, R.Murthy.Oracle8i-the XMLenabled data management system. *Data Engineering, Proceedings. IEEE Int'l Conf. 16th International Conference*, 2000, pp.561-568.
- [6]. UML For W3C XML Schema Design, [Online Available] http://www.xml.com/pub/a/2002/08/07/wxs_uml.html#XMI
- [7]. Weibel, S., Godby, J., & Miller, E. .OCLC/NCSA metadata workshop report, 1995.
- [8]. 陳亞寧、陳淑君、沈漢聰、鍾豐謙,後設資料系統的需求評選與發展設計，第二屆數位典藏技術研討會,台北：中央研究院資訊科學研究所, 2003.
- [9]. 陳昭珍,電子資源長久保存,佛教圖書館館訊,Vol. 26, No.2, 2001.
- [10]. 陳雪華、項潔、鄭惇方,數位典藏在數位內容產業之應用加值,數位內容創意加值研討會, 2003,台北：國立台灣大學.
- [11]. 蔡順慈、林昱伍, 數位典藏工作流程規劃-以華岡博物館為例,數位典藏工作流程與品質管理研討會, 2004.