

行政院國家科學委員會補助專題研究計畫成果報告

數位博物館專案— 負責數位博物館權威控制系統之設計

計畫類別：☐ 個別型計畫 ☒ 整合型計畫

計畫編號：NSC 89-2750-P-002-014-

執行期間：89年3月1日至91年2月28日

計畫主持人：陳光華

本成果報告包括以下應繳交之附件：

- ☐ 赴國外出差或研習心得報告一份
- ☐ 赴大陸地區出差或研習心得報告一份
- ☒ 出席國際學術會議心得報告及發表之論文各一份
- ☐ 國際合作研究計畫國外研究報告書一份

執行單位：國立台灣大學圖書資訊學系

中華民國九十一年二月二十八日

目次

壹、緒論.....	3
一、研究背景.....	3
二、研究目標.....	4
三、研究方法.....	6
四、研究進行步驟.....	6
貳、文獻分析：數位圖書館與博物館的發展.....	8
一、國科會「數位博物館專案計畫」.....	9
二、國科會「國家數位典藏數位化計畫」.....	12
三、文化建設委員會「全國文化資料庫」.....	14
四、數位化資源組織相關計畫.....	17
五、結語.....	23
參、問題研討與分析.....	28
一、數位圖書館/博物館.....	28
二、數位圖書館/博物館的資訊組織.....	32
三、XML/Metadata 輸入、交換及管理系統的設計.....	33
四、數位博物館的權威控制.....	34
五、數位博物館的整合查詢.....	35
肆、中文索引典的研究與應用.....	38
一、數位圖書館與數位博物館.....	38
二、索引典.....	39
三、查詢問句擴展.....	45
四、中文索引典的應用.....	50
伍、XML/Metalogy 管理系統之權威控制系統建置.....	59
一、數位圖書館系統架構.....	59
二、Metadata 之意義與功能.....	60
三、Metadata 的語法—SGML 與 XML.....	64
四、XML/Metadata 系統發展實例.....	71
參考文獻.....	89

圖 表 目 次

圖 1：權威控制子系統與其他子系統之交互運作模式.....	5
圖 2：全國文化資料庫之徵集方式.....	16
圖 3：Us-EU 研究小組對於數位圖書館/博物館的研究範疇.....	29
圖 4：XML/Metadata 系統架構.....	34
圖 5：數位博物館的權威控制.....	35
圖 6：HTTP/Z39.50 gateway 之系統架構.....	36
圖 7：詞彙的語義關係.....	40
圖 8：“Education” 主題之下可進一步選擇的層面.....	43
圖 9：供使用者選擇的敘述語.....	43
圖 10：經使用者與系統互動後的檢索條件.....	44
圖 11：查詢問句擴展程序.....	46
圖 12：數位圖書館系統架構.....	59
圖 13：Metadata 的三個問題.....	63
圖 14：Metalogy 系統架構圖.....	73
圖 15：資料庫檔案關連圖（未包含權威控制檔）.....	75
圖 16：編目參數檔、資料檔與權威檔之關係.....	78
圖 17：編目資料流程圖.....	79
圖 18：查詢功能資料流程圖.....	82
表 1：XML 與 HTML 在特性上主要的不同.....	66
表 2：XML 的特別保留字元.....	69
表 3：XML 與 SGML 的相同點.....	70

壹、緒論

台灣地區有許多學術研究單位、圖書館、博物館等機構珍藏史籍文獻與各種器物。這些為數甚多的珍藏，為避免資源的加速損壞，無法開放讓大眾參觀。有些或者可以開放參觀，但因受限於時空的限制，許多人無法一睹其藏品的內涵。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際網路，使其具有恆久的生命，開放給更廣泛的使用群眾，是首要的工作。其次，目前網路上的資源處於無序的狀態，因此第二個重要的工作就是必須針對不同資源的特性，有效地進行組織與整理的工作。

一、研究背景

我國國科會自八十七年開始進行「數位博物館」專案計畫。有數位圖書資訊學界的教授們，共同執行該專案計畫之下的一個分項計畫，名稱為：「資源組織與檢索之規範」(Resources Organization and Searching Specification, 簡稱 ROSS)。ROSS 亦於 1999 年參與 CIMI DC 第二階段的實驗計畫。

ROSS 之研究目標涵蓋在中文環境數位圖書館／博物館中，有關於資訊組織與檢索各個層面的重要議題；包括資料的儲存與管理系統的設計、使用者資訊需求及查詢資料的方式、各系統間整合檢索等。

ROSS 詮釋資料研究群在八十七年數位博物館專案計畫中已獲得相當多的研究成果：包括：1. MICI-DC；2. MICI-DC XML DTD；3. Metalogy (XML/metadata 輸入及資料庫管理) 雛形系統；4. MICI Z39.50 Profile；5. Z39.50 Client / Server 雛形系統；6. 權威檔建立及瀏覽雛形系統等。詳細說明請參見本章以後的資料。

雖然 ROSS 詮釋資料研究群的努力有一初步的成果，但是尚有許多問題有待解決，例如需要進行使用者資訊需求研究，以便決定何者為檢索點 (access

points)。另外，亦須探討與決定有哪些欄位為必備欄，有哪些為非必備欄；或是決定哪些是屬於核心欄位，以做為不同系統之間資料檢索交換之用。另外，ROSS所發展的詮釋資料管理系統 Metalogy，Z39.50 Client / Server 系統，以及權威檔建立及瀏覽系統也還有部分尚未完成。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：

1. 繼續發展詮釋資料管理系統Metalogy。
2. 繼續發展Z39.50 Client / Server系統。
3. 繼續發展權威檔建立及瀏覽系統。
4. 繼續探討現有國外幾個重要詮釋資料格式發展之觀察與追蹤。
5. 發展MICI的使用者著錄手冊。
6. 探討多語言詮釋資料資訊查詢檢索的可行性，以便與國際間作交換。
7. 邀請相關館藏單位參與實驗計劃使用MICI著錄其藏品，並定期討論以做為改良詮釋資料格式的參考。

未來，本研究群期望所規劃設計之詮釋資料，不僅可以符合國科會「數位博物館」專案計劃中藏品的特性，還可以與國際上其他國家數位圖書館／博物館之詮釋資料相容。更期望所發展的詮釋資料管理系統 Metalogy，可提供相關館藏單位實際輸入資料之用。

二、研究目標

本計畫的目標涵蓋在中文環境數位博物館/圖書館中，有關於資訊組織與檢索各個層面的重要議題，包括：資料的儲存與管理系統的設計、使用者資訊需求及查詢資料的方式、各系統間整合等。本計劃第二年目標不僅在於支援國科會「數位博物館專案計畫」主題計畫中有關於資訊組織與檢索相關的部分，發展重點在於制定國內中文數位博物館/圖書館環境中資源組織與檢索之相關規範，以及發

展一套能夠實做的系統。這些規範要考量能夠與國際相關標準相容的問題，使中文資訊能被國外之整合檢索軟體檢索，並促進國內數位博物館/圖書館系統國際化的重要管道。通透性的整合檢索是電子資訊檢索系統中不可或缺的一項功能，也是中文數位博物館/圖書館系統國際化相當重要的課題。因此，本計畫積極探討並開發符合國際標準之整合檢索系統，以免讓國內之中文數位博物館/圖書館成為閉關自守的系統。具體而言，本子計畫的目標列述於下：

子計畫二：負責數位博物館權威控制系統之設計

1. 探討索引典與標題表在數位博物館的應用情形；
2. 分析數位博物館系統中各子系統之交互運作性及權威控制子系統之角色；
3. 探討數位博物館系統中權威檔案、索引典、標題表之呈現模式；
4. 發展數位博物館系統中權威檔案、索引典、標題表之組織模式；
5. 訂定數位博物館系統中權威子系統之各項規格；
6. 建立數位博物館系統中權威控制之離型系統；

權威控制子系統與其他子系統之交互運作模式如下圖所示。

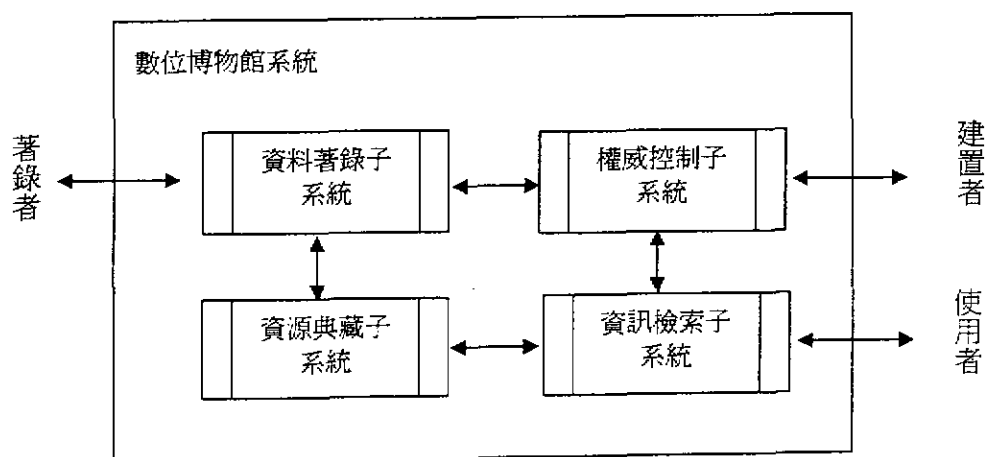


圖 1：權威控制子系統與其他子系統之交互運作模式

三、研究方法

子計畫二：負責數位博物館權威控制系統之設計

本子計畫執行方法與實施步驟包括：文獻分析、系統分析、系統實作、系統評估等。分別敘述如下：

1. 文獻分析法：

蒐集國內外權威檔、索引典、標題表相關論文，實際應用的研究文獻，以及相關的國際標準。

2. 系統分析法：

確認目前線上索引典、標題表的展現模式，分析數位博物館系統中各子系統的互動模式，訂定權威控制子系統之功能，設定權威檔案結構，擬定系統實作流程。

3. 系統實作法：

按照系統實作流程，使用C++與Java語言，系統平台為Windows NT與Windows 98，建置雛型系統。

4. 系統評估法：

評估雛型系統之系統功能，與其他子系統之互動功能。

四、研究進行步驟

本計畫的研究目標為：在中文環境的數位圖書館與博物館中，建立資源組織與檢索的規範以及實作系統。其研究進行步驟如下：

子計畫二：負責數位博物館權威控制系統之設計

1. 訂定數位博物館權威控制系統之規格。
2. 訂定系統權威檔之呈現模式、組織格式、瀏覽模式。
3. 訂定數位博物館各子系統之互動模式。

4. 確認權威控制系統之細部功能。
5. 建置數位博物館權威控制子系統。
6. 訓練參與研究計畫之助理，培養數位博物館相關研究之人才。
7. 檢討雛型系統、撰寫報告、並發表初期研究成果。

貳、文獻分析：數位圖書館與博物館的發展

隨著網際網路全球性的快速發展，各國相繼努力推廣網際網路的普及應用，整個網路使用人口在近幾年驟增。台灣網路人口由 1996 年 4 月之 40 萬，成長至 2001 年 6 月止之 721 萬，有 18 倍的成長。(註 1) 除了上網人口的顯著增加外，政府、工商企業等各單位，截至 2001 年春季，所建置的網際網路連網主機已超過 130 萬個，較 1996 年的 4000 多個，成長 325 倍之多。(註 2) 各種網路資訊應用與服務誠然已蓬勃發展，這意謂著資訊網路化的社會將蔚然成型，使得人類溝通的方式以及知識訊息表達與傳遞的方式有巨幅的轉變。過去人類所創的文化資產與各類知識，可以藉科技加以保存、整理、傳播與利用，使知識與文化資源得以全球共享，繼而加速文明的進步。

目前，數位圖書館／博物館的研究隨著網際網路的發展相當受到世人的重視，各先進國家無不熱烈支持有關的各項計畫，如聯合國的 Memory of the World Programme、美國國會圖書館(Library of Congress)的 American Memory、美國國家科學基金會(National Science Foundation)的數位圖書館先導計畫一、二階段 (Digital Library Initiative & phase 2)、教廷梵諦岡的珍貴手稿數位化、法國羅浮宮的繪畫數位化、澳洲的澳洲線上博物／美術館計畫 (Australian Museums & Galleries On-line)、澳洲文化網 (Australia's Cultural Network)、日本的次世代數位典藏系統研究與發展專案 (Generation Digital Library System Research and Development Project)、日本的全球數位博物館計畫 (Global Digital Museum) 等。(註 3)

我國有悠久的歷史文化，累積豐富的文化資產，是世界的瑰寶；而關於台灣本土的珍貴文獻與歷史文物，許多機構與學術研究單位有豐富的典藏。台灣地區目前有許多機構從事珍貴藏品數位化的工作，包括：國家圖書館、故宮博物院、歷史博物館、自然科學博物館、中央研究院、臺灣大學、文化建設委員會、文化中心等。本文介紹近年來台灣地區主要的數位化計畫之發展，包括數位博物館、國家數位典藏、以及全國文化資料庫等，並特別針對資源組織之部分加以說明。

一、國科會「數位博物館專案計畫」

國家科學委員會（簡稱國科會）為了加強人文社會科學的研發以及科學教育工作，從 1998 年 5 月開始推動「迎向新千禧—以人文關懷為主軸的跨世紀科技發展」方案。「數位博物館」專案即為此方案中的計畫之一，其主要目標為：整合建置一個適合國情並具有本土特色的「數位博物館」，以發展教育性網際網路內涵。（註 4）藉由無遠弗屆的網際網路建立並推動文化、藝術、科技等教育性網站內涵的典範，使一般大眾得以不受時空限制，隨時上網檢索或瀏覽並利用其資訊，進而豐富人民生活的內涵與享受終身學習的樂趣。（註 5）此外，更希望藉由數位典藏的推動，刺激多媒體數位典藏技術與產業之發展。

「數位博物館」專案計畫自 1998 年 9 月推動以來，目前已進入第三階段。國科會在第一年首先邀集中央研究院、臺灣大學、清華大學等已進行數位典藏與相關技術研究的專家學者，共同建立一個「數位博物館」的推動與合作機制。該專案的推動大致區分為主題計畫以及技術支援等兩大類。另外，還有「推廣教育計畫」，係培養數位典藏種子人才，並將主題計畫與技術支援計畫推廣給各界使用。

（一）第一期（1998.09－1999.10）（註 6）

1. 主題計畫

從內容來看，第一階段的主題計畫包括了本土風情和傳統文化兩方面。在本土風情方面，有二個綜合性質的人文主題計畫，即：淡水河溯源（台灣大學圖書館）和臺灣原住民—平埔族群（中央研究院民族學研究所）。在自然科學與自然生態方面則有：蝴蝶生態面面觀（暨南大學資訊工程系與自然科學博物館）、台灣的魚類（中央研究院動物研究所）和台灣本土的植物（中央研究院植物研究所）。在傳統文化方面有：傳統思想與文學（四書、老莊、唐詩）（中央研究院語言學研究所籌備處）、不朽的殿堂—漢代的墓葬與文化（中央研究院歷史語言研究所）、火器與明清戰爭（清華大學歷史系）等三個計畫。

2. 技術支援計畫

在共通技術支援的系統方面，第一階段已著手建立的有：關於時間、空間和語言文字等共同的座標體系，此即人文與自然資源地圖（中央研究院計算機中心）和搜文解字—語文知識網路（中央研究院語言學研究所籌備處）兩個計畫的主要內涵。此外，還有一些共同的資訊技術項目，包括：資源組織與檢索之規範（台灣大學圖書資訊學系）、系統評估（台灣大學圖書資訊學系），以及數位典藏系統先導計畫（中央研究院資訊科學研究所）等。資源組織與檢索之規範為主題計畫建立中文詮釋資料（metadata）交換格式、索引典以及檢索與搜尋等規範，使各主題系統具有國際化的透通性。系統評估以淡水河溯源計畫為例，建立主題系統建置過程中的各項評估規範與準則，促使各主題系統具有較高的品質與傳播效果。

3. 推廣教育計畫

數位博物館專案推廣辦公室（台灣大學圖書館）成立於1998年8月，其目的在於培育數位典藏種子人才，並將計畫成果擴散而廣為社會各界運用。推廣辦公室透過研討會、教師網路素養訓練班、專業訓練課程、媒體專文介紹、電子報傳送等方式，加強台灣地區典藏單位與產業界對數位圖書館與博物館的認識，提昇民眾的興趣，充實中小學教師使用網路資源能力，並培養建置數位圖書館與博物館的專業人才。

（二）第二期（1999.09－2000.10）（註7）

1. 主題計畫

本階段開放並擴大參與的單位，在近九十件申請的主題計畫中核准十二件，包括：1.故宮文物之美（故宮博物院）、2.玄奘西域行（台灣大學哲學系）、3.淡水河溯源數位博物館（台灣大學圖書館）、4.台灣民間藝術家數位博物館之建置--以楊英風數位藝術博物館為例（交通大學）、5.台灣老照片數位博物館（藝術學院）6.台灣建築史（義守大學）、7.生命科學數位博物館--人體奧秘展覽館（台北醫學院）、8.臺灣原住民--平埔族群數位圖書館與博物館（中央研究院民族學研究所）、9.中小學語言文學知識網路（中央研究院語言學研究所籌備處）、10.尋回台灣本土的

淡水魚類（中央研究院動物研究所）、11.中醫藥、針灸虛擬數位博物館（中國醫藥學院）、12.蘭嶼生物/文化多樣性數位博物館（暨南大學資訊工程系）等。以上十二件計畫當中，有四件計畫（編號3, 8, 9, 10）係延續第一年的計畫。

2. 技術支援計畫

本階段之技術支援計畫則減為兩件：資源組織與檢索之規範與系統實作（台灣大學圖書資訊學系）以及2.數位博物館影像版權資訊植入技術與軟體之開發（交通大學資訊工程學系）。

國科會「數位博物館專案」計畫預期效益有三：（註8）1.奠定我國未來網路世紀文化內涵建置的基礎，2.拓展電腦多媒體技術深層應用，作為多媒體文化產業長遠發展的基礎。3.藉網路媒體宣揚、發揮我國既有文化的價值。

（三）第三期（2001.05－進行中）（註9）

1. 主題計畫

本階段是以主題計畫為主，其內容包括「語言文學與宗教藝術類」、「民俗文化與歷史文物類」、「數學生物類」、及「建築與地理類」等四類。

- 語言文學與宗教藝術類：1.台灣民間藝術家數位藝術博物館 -- 楊英風、蘇森墉、高棧（交通大學圖書館）、2.「荔鏡姻，河洛源」 -- 閩南語第一名著（荔鏡記）的數位博物館（搜文解字 III）（元智大學中國語文學系）、3.數位博物館計畫：玄奘西域行（II）（台灣大學哲學系）、4.台灣老照片數位博物館（藝術學院傳統藝術研究中心）。
- 民俗文化與歷史文物類：1.商王大墓重現 -- 本世紀美術考古大發現（中央研究院歷史語言研究所）、2.故宮文物之美系列（故宮博物院）、3.織品服飾數位博物館 -- 以苗族為例（輔仁大學織品服裝所）、4.台灣社會人文電子影音數位博物館（藝術學院科藝技術研究中心）。

- **數學生物類：**1.生物／文化多樣性數位博物館之建構(II) -- 阿里山山脈與鄒文化數位博物館 (自然科學博物館人類學組)、2.生命科學數位博物館 -- 生命密碼館 (台北醫學大學資訊研究所)、3.台灣昆蟲與ㄇㄣˊ類資源數位館之建立 (農委會農業試驗所應用動物系)、4.昆蟲數位化博物館之建立 -- 蟲蟲總動員 (台灣大學昆蟲學系)、5.台灣物理治療數位博物館 (台灣大學醫學院物理治療學系)。
- **建築與地理類：**1.福爾摩沙文化生態地圖 (台灣大學地理環境資源學系)、2.數位博物館專案主題 -- 「台灣古蹟巡禮」 (中國技術學院建築工程系)、3.台灣建築史 (義守大學)。

綜而言之，國科會「數位博物館專案計畫」之目標為：在網路上開發有意義、高品質的主題系統及技術，建立文化藝術性、科學教育性的內涵教材；建立國內數位博物館合作發展機制與環境。推廣數位博物館與培養相關人才與產業。擴大網路內涵對社會及教育的良性影響。

二、國科會「國家數位典藏數位化計畫」

1999 年 7 月，行政院第九次「電子、通訊、資訊策略會議」(SRB) 通過了「國家典藏數位化計畫」，並將此案分由國科會協調執行。本計畫於 2001 年 1 月正式展開，其目標是將故宮博物院、國家圖書館、國立歷史博物館、台灣省文獻委員會、自然科學博物館、台灣大學和中央研究院等七個單位珍貴的重要的文物典藏數位化，建立國家數位典藏，以保存文化資產、建構公共資訊系統，促使精緻文化普及化、大眾化，資訊科技和人文融合，並推動產業和經濟發展。此專案計畫九十年度的計畫內容摘要如下：(註 10)

(一) 故宮文物數位典藏系統之規劃與建置

主要目標為建立國寶與重要古物級文物較完整之數位典藏資料庫，包括精華的器物和書畫，以及清代檔案之專業攝影及數位化建檔，並完成文物資料庫建置。

(二) 國家圖書館國家典藏數位化

計畫內容涵蓋:1.古籍文獻，以明人詩文集為優先，並將古籍中之附圖版畫及金石拓片亦數位化。2.館藏及地方之相關文獻以及 3.期刊報紙。

(三) 歷史博物館國家歷史文物數位典藏資料庫發展計畫

主要目標將該館過去累積四十餘年珍貴資料全面數位化外，亦著重將進行中各類資訊、知識收入多媒體資料庫管理使用。

(四) 台灣省文獻委員會典藏日據與光復初期史料數位化

計畫內容涵蓋典藏日據與光復初期史料數位化，包括檔案文書原始手稿的分類編目、影像掃描、索引建檔、檢索系統的開發與建置等任務。

(五) 自然科學博物館國家典藏數位化計畫

主要目標將該館豐富且具代表性及獨特性之自然史及文化史典藏資源與予數位化及網路化。

(六) 台灣大學典藏數位化計畫

計畫內容涵蓋歷史文獻、考古、地質、動物、植物、昆蟲等，其中多與臺灣本土有關。擬優先將有關臺灣文獻文物的典藏品（如，淡新檔案、岸裡大社文書、伊能手稿等）予以數位化，進行影像掃描及全文建檔，並建置資料庫以供學者查詢利用。

(七) 中央研究院國家典藏數位化計畫

計畫內容涵蓋魚類、貝類、植物、地圖、台灣原住民風俗、台灣原住民語言、台灣地方檔案、近代經濟外交文獻、中原考古文物、金石拓片、古文書、文獻、善本圖籍、語料、及數位典藏技術等。

以上這些計畫在 91 年已經進一步提升為：「國家數位典藏」國家型計畫，該計畫規劃在組織架構上設有內容發展、應用服務、訓練推廣、技術研發、行政支援等五組。在內容發展主題計畫方面，主要包括：中央研究院國家數位典藏數位化計畫、典藏日據與光復初期史料數位化、國家圖書館國家典藏數位化計畫、台灣大學典藏數位化、國家歷史文物數位典藏計畫、故宮文物數位典藏系統之規劃與建置、自然科學博物館國家典藏數位化計畫、國史館典藏國家檔案與總統文物數位化、文化藝術數位博物館。除了今年「國家數位典藏數位化計畫」中七個典

藏單位之外，另包括國史館與文建會兩個單位。

本國家型計畫的目標在於將國家重要的文物典藏數位化，建立國家數位典藏；之後再進一步地以國家數位典藏促進臺灣地區人文與社會、產業與經濟的發展。

三、文化建設委員會「全國文化資料庫」

文化建設委員會（簡稱文建會）「全國文化資料庫」之目標在於規畫一個數位化國家文化資料庫之架構與建置規範。該資料庫期望能具備的功能有：（註 11）

1. 提供全民參與文化建設的管道，蒐集並整理台灣文化資產
2. 整合地展現全民的文化資產
3. 提供文化資料累積、共享與活化利用的機制

雖然「全國文化資料庫」從 2001 年春天開始執行，其實在 1999 年即開始進行「地方文獻數位化計畫」，算是其前身。該計畫為台灣省政府文化處（精省之後改名為文建會中部辦公室）「台灣省公共圖書館資訊網路輔導諮詢委員會」地方文獻數位化模式及相關標準研究小組於 1999 年進行「地方文獻數位化模式及相關標準研究」。根據研究結果顯示，各相關機構對於地方文獻蒐藏的資料類型，以照片最為普遍，其次分別為口述歷史、文物、方志、族譜、歌謠/諺語、私文書、報告書、詩文、公文書、碑碣、繪畫、見聞、海報傳單、書簡、戲劇、戰記、競選資料、日記、計文、備忘錄等。由於數位化的工作相當耗時費力且所費不貲，而地方文獻相關資料又如此多，必須訂出數位化的優先順序。因此研究小組以文化中心之特色館之蒐藏主題，訂為各文化局在此計畫內要進行數位化的範圍。又因經費有限，第一年（2000 年）被選定參與的文化局有：宜蘭縣（台灣戲劇）、台北縣（現代陶瓷）、新竹縣（客家文物）、台中縣（編織工藝）、高雄縣（皮影戲劇）及屏東縣（排灣族雕刻藝術）等。（註 12）

「全國文化資料庫」開始執行之後，含括前述「地方文獻數位化計畫」的藏品，並闢增其藏品內容，所包括的資料類型有：音樂資料、美術資料、戲劇資料、舞蹈資料、傳統藝術資料、文學資料、電影資料、建築資料等八大類。其原始資料的來源有兩種：一為已經由中央與地方文化機關／單位收藏的公藏文化資料，二為尚藏於民間的文化資料：(註 13)

(一) 公藏資料：

公藏資料指已經由公立博物館、美術館、檔案館、文物館，或其他各種公立機關收藏保存之資料。這些公立機關如下：文建會轄下機關與單位包括：台灣工藝研究所、臺灣交響樂團、傳統藝術中心籌備處、文化資產中心籌備處、藝術村籌備處、民族音樂中心籌備處、臺灣博物館、台中圖書館、臺灣美術館、臺灣歷史博物館籌備處等。縣級文化單位則包括縣市文化局與縣市文化中心等。

(二) 民間資料：

民間之藏有資料者，成員相當複雜，有創作者（通常也是收藏家）、個人之專業或業餘收藏家、公共性質之基金會或私人博物館、美術館等。另外還有文史工作室、藝術工作室、文化藝術社團、文化藝術產業等。

在資料徵集方面分為主動徵集與接受申請兩種方式。主動徵集為由文建會本會及轄下各機關／單位，主動接洽民間之資料擁有者，獲其同意將其藏品提供數位化，存入資料庫。接受申請則為：由文建會及轄下各機關／單位訂定辦法，鼓勵民間之資料擁有者提出申請，將其資料數位化，匯入資料庫。匯入時的接收承受窗口，以文建會轄下之各博物館、美術館、中心等機關，及縣級文化單位為主，文建會之各業務處為輔。全國文化資料庫的徵集方式請參見下圖。

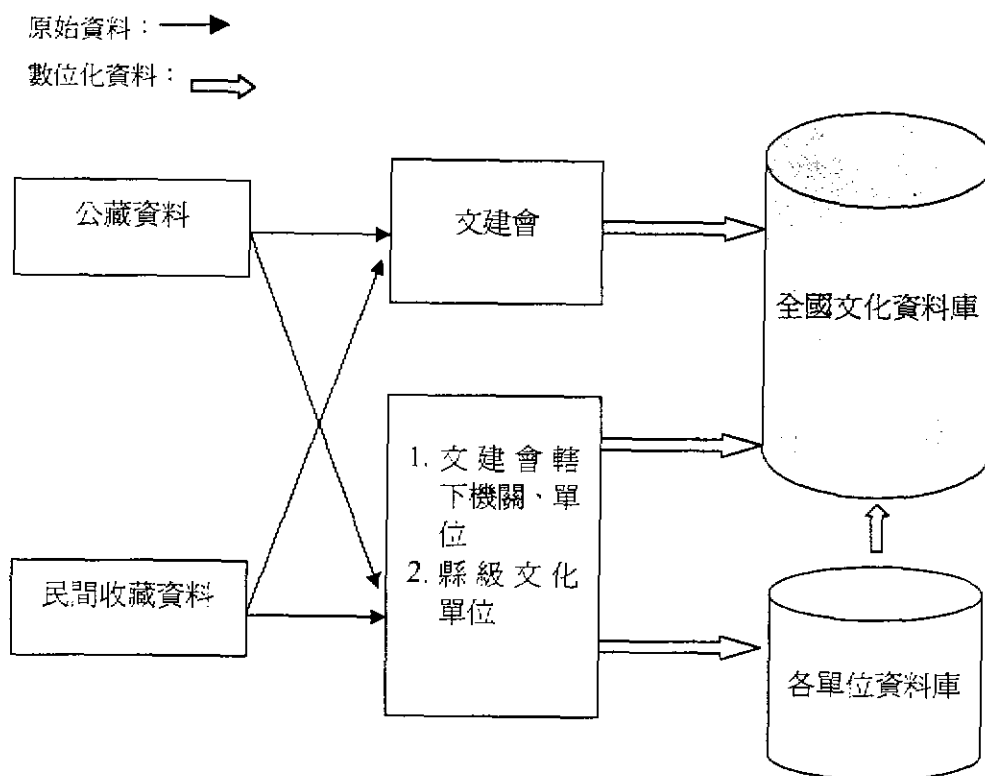


圖 2：全國文化資料庫之徵集方式（註 14）

全國文化資料庫之基本架構包括中央整合資料庫與分散度藏資料庫：（註 15）

1. 中央整合資料庫（文建會本部）：全國性的管理系統，收集各地送來的數位化文化資料，作為國家數位文化資料保存中心，並提供未來各種不同的用途。收藏由分散度藏資料庫所送來的數位化文化資料，並供全國民眾檢索使用。資料庫亦可接受來自於網際網路的查詢請求，透過資料庫的查詢機制，提供外界各種不同的文化資料相關資訊。
2. 分散度藏資料庫：分散度藏資料庫是由文建會轄下各機關／單位，及縣級文化單位，建置各類型文化資料的數位資料庫，將文化資料予以適當的數位化處理，並賦予詮釋資料，形成一個具有強烈語意描述的數位化文物物件。並透過網際網路傳遞的形式，將各式文物的數位化資料上載到中央整合資料庫。

目前許多部會也有許多數位化的計畫，國內數位化的趨勢已蔚為風潮。除了前述國科會的『數位化博物館』專案計畫、『國家典藏數位化』計畫以及文建會的「全國文化資料庫」等三個計畫外，教育部也有一些大型計畫與數位典藏有關，如：資訊教育總藍圖、卓越計畫、電腦輔助教學、遠距教學、海外華語教學。另外，研考會的電子化政府、內政部的國土地理資訊系統、經濟部的電子商務等許多計畫都和數位化有關。(註 16)

四、數位化資源組織相關計畫

隨著大規模數位化計畫的執行，資訊累積的速度越來越快，從管理者角度觀之，必須適當地整理並組織資訊，使其處於有序的狀態，才能提供使用者有效的資訊檢索。如何將數位網路資訊資源納入資訊組織領域內，以及有效達成精確、快速的資訊檢索就更顯得重要。在台灣地區，近年來有許多學者專家積極投入數位資訊組織相關的研究。茲將數位化資源組織相關計畫，包括中文詮釋資料 (Metadata) 的研發與管理系統實作等方面的活動，依照成立的先後順序分述於下：

(一) 「資源組織與檢索之規範」研究計畫

1998 年筆者與一些圖書資訊學領域學者共同合作執行一項名為「資源組織與檢索之規範」(Resources Organization and Searching Specification, 簡稱 ROSS, <http://ross.lis.ntu.edu.tw>) 專案計畫，作為國科會「數位博物館」專案計畫下的一個分項計畫，其研究目標涵蓋在中文環境數位圖書館／博物館中，有關於資訊組織與檢索各個層面的重要議題，包括：資料的儲存與管理系統的設計、使用者資訊需求及查詢資料的方式、各系統間整合等。ROSS 計畫主要的工作為支援國科會「數位博物館」專案計畫各主題系統，其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規範。(註 17)

在國科會的「數位博物館」專案計畫開始之前，本研究群於 1997 年 3 月在「臺大電子圖書館與博物館」(National Taiwan University Digital

Library/Museum，簡稱 NTUDL/M）計畫中即成立了「詮釋資料研究群」，著手於中文資訊詮釋資料交換（Metadata Interchange for Chinese Information，簡稱 MICI）之相關研究。詮釋資訊研究群也於 1998 年 11 月開始為國科會「數位博物館」專案計畫展開各類型藏品（器物、古地圖、圖像照片、蝴蝶標本等資料）詮釋資料的制定。（註 18）

ROSS 研究計畫第一年（1999 年）之總計畫名稱為「數位博物館專案計畫：資源組織與檢索之規範」，其分為五個子計畫，分別進行資源組織與檢索規範的各個層面之相關研究，包括：電子資源之組織與中文資源著錄格式的制訂、使用者資訊尋求行為分析與便易使用環境之建立、索引與架構及權威資料庫之建置、資訊檢索及資料搜尋服務系統的設計、以及數位博物館/圖書館整合檢索。其工作主要是支援「淡水河溯源」和「蝴蝶生態面面觀」等兩個計畫的詮釋資料之制定；而第二年（2000 年）之總計畫名稱為「數位博物館資源組織與檢索之規範與系統實作」，其分為三個子計畫，分別是：中文電子資源著錄規則暨格式之制定、數位博物館權威控制系統之設計、以及 XML/Metadata 輸入、交換及管理暨整合檢索系統的開發等。ROSS 計畫第二年主要任務是發展一個能使各種詮釋資料並存的管理系統 - Metalogy，以提供有需要的各個主題計畫使用。（註 19）

目前 ROSS 研究計畫已經完成，其主要的研究成果包括：中文詮釋資源交換格式（MICI-DC）、MICI-DC 著錄手冊、XML/Metadata 管理系統（Metalogy）、以及 Metalogy 安裝手冊等。ROSS 已將這些成果載於《數位圖書館 XML/Metadata 管理系統》書中。（註 20）茲將這些成果簡述於下：（註 21）

中文詮釋資料交換格式（MICI）乃以 Dublin Core 十五個基本欄位為主要的架構。但是為了描述文化遺產豐富的屬性、更精確表達描述藏品的語意，MICI 按照藏品類型特色，在相關欄位下，搭配欄位修飾語（element qualifier）的使用，不僅可以擴大其應用的範圍，同時也具有國際性。我們將以 DC 核心集為基礎並加上自行定義的欄位修飾語的詮釋資料格式稱為 MICI-DC。目前 MICI-DC 可處理的資料類型包括：古文書、地圖、圖像/

照片資料、器物以及故宮的書畫、器物與文獻（佛經、佛經圖繪）。除採用 DC 制訂之修飾語外，也根據各典藏單位之需求自行定義相關修飾語，使用者可自行選擇 DC15 個欄位與修飾語，並視需求調整欄位順序，在遵循國際標準外，同時提供最大彈性以符合使用者的需求。為讓使用者更易於著錄 MICI-DC，ROSS 製作了 MICI-DC 著錄手冊，根據 15 個欄位及其修飾語做語意上的說明，並提供相關範例，讓一般使用者可以自行選擇彈性運用。

由於 XML 已成為網路界及資料庫界積極支持的語言。因此，當 ROSS 要設計一個詮釋資料管理系統時，也決定採用這種語言，做為資料庫間交換資料的主要依據。同時，由於目前已發展出來的詮釋資料格式相當多，為了能表達各類型各主題資料之屬性，尚未發展詮釋資料格式的領域可能還會發展出該領域的詮釋資料格式，所以一個有彈性的詮釋資料管理系統，不能只是針對一種詮釋資料格式來開發，而應該能讓使用單位自己決定要用那一種格式。所以，開發一個 XML/Metadata 的通用性系統正好可以滿足這種需求，這也是 Metalogy 這個系統最主要的特色。Metalogy 系統可以用來建立數位博物館、數位圖書館、數位檔案館或任何主題資料之資料庫。它具有利用 DTD 建立資料庫、編輯詮釋資料、編輯權威檔(或索引典)、查詢(含 Window 介面及 Web 介面)、及轉入及轉出 XML 記錄等功能。為了便於使用者利用本系統，ROSS 也製作了 Metalogy 操作手冊，以圖示法詳細說明本系統的安裝、設定、使用及操作方法。

另外，為使 Metalogy 系統能讓大眾廣為參考及運用，國科會數位博物館專案推廣辦公室於 2001 年 7 月 4 日舉辦「Metadata 通用系統 (Metalogy) 技術開放發表會」，並邀請各典藏單位以及軟體發展廠商參與。

(二) 中央研究院「Metadata 工作小組」(註 22)(註 23)(註 24)

中央研究院擁有豐富的各種典藏品，無論在形式與學科上都橫跨眾多的專門學科領域，但整體言該機構是以文化、歷史傳承的特殊典藏品作為數位圖書館博物館計畫的主要範圍。為達有效應用詮釋資料與發揮數位圖書館博物館的功能，中央研究院自 1998 年 11 月正式成立「Metadata 工作小

組」，負責該院數位典藏計畫（Digital Archive Project）中有關詮釋資料的規劃、設計與實施等各項事宜。研究目的是要先整合中研院計算中心與資訊所各個相關實驗室的既有成果，參考國際研究趨勢，並且與研究同仁的密切溝通配合，建立一套符合全世界華文使用者習性的多媒體數位典藏的整合管理與服務環境，提供收集、校對、後製、典藏、傳播、呈現、系統安全與管理、以及使用者服務等各項數位典藏系統的基礎功能。中研院 Metadata 工作小組除了以中研院數位典藏計畫為主外，同時也支援其他相關計畫的詮釋資料，如「臺灣地震數位知識庫」，以及國家科學委員會委任執行的各項「數位圖書博物館計畫」。（註 22，23）該 Metadata 工作小組之主要任務有三：（註 24）

- 制定符合該院進行的數位博物館專案計畫需求的詮釋資料格式；
- 以該院詮釋資料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詮釋資料格式與元素；
- 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）。

此外，該工作小組立下要建構一套兼容臺灣本土典藏特性、又能兼顧該院學術研究需求、並能符合網路資源國際化的詮釋資料的目標，因此規劃了以下八項原則做為設計該院詮釋資料時之依據：（註 25）

- 符合國際標準要求，包括詮釋資料格式、網路協定、檢索等；
- 依據本土需求採用既有標準進行修改，不再另行重新研發；
- 資料的建立、展現、檢索、詮釋與串聯關係的互動要容易使用；
- 具備多語文能力；
- 因應不同學科領域需要採用不同的詮釋資料格式；
- 採用不同的詮釋資料格式時，利用 XML 建置詮釋資料基礎骨幹與結構來容納、整合，包括與現有圖書館自動化系統的結合，以利一次查詢所有的館藏資源；

- 設計系統管理機制功能，以因應不同的需求，包括取用限制、智慧財產權的保護；
- 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。

為達到前述的任務與目標，該工作小組以 1).現有文獻研析及詮釋資料標準格式與元素分析；2).了解院中各單位資料庫系統建置情形；3).實際訪談了解資料特性與使用者需求；4).制定適合該院需求的詮釋資料；5).以「臺灣原住民－平埔族群」計畫為首例，進行詮釋資料的實作與修正；6).調查臺灣本土典藏資料特性及系統建置的實際情況；8).完成系統設計與運作雛型；9).舉辦公聽會徵求意見，以完成系統設計與實作等九個步驟來進行中央研究院詮釋資料格式之研究與設計。(註 26)

(三) 國家圖書館「研擬圖書資訊相關技術規範」計畫

由於詮釋資料議題的重要性，國家圖書館 2000 年成立「Metadata 研究小組」，擬凝聚臺灣地區詮釋資料有興趣的單位與學者專家，整合國內已發展之詮釋資料。該小組已於 2000 年 8 月 17 日召開第一次會議，2000 年 11 月 8 日第二次會議。第一次會議決議由國家圖書館設計一個詮釋資料首頁，提供國內發展詮釋資料之單位一個公告的園地，同時也藉此讓各界有針對該詮釋資料提出建議的機會，做為修訂之參考。並將修訂完成之詮釋資料，由國家圖書館加以匯整印製出來，供國內各界使用時的參考。國內發展之詮釋資料，應積極在 DCMI (Dublin Core Metadata Initiative) 登記，以取得國際的認同。該研究小組之任務包括：探討國內外詮釋資料發展現況；積極加入國內外詮釋資料研究組織及相關計畫（如 DC 會議及 CORC 計畫）；發展各主題領域或資料類型之詮釋資料；研究 MARC 與詮釋資料轉換之語法、語意等相關標準；發展詮釋資料管理系統；詮釋資料首頁的維護與公告；善本古籍詮釋資料的設計；MARC DTD 的設計；CMARC DTD 轉成 DC-qualifier 的設計等。(註 27)

此外，國家圖書館也收集了國內各單位已發展出來的詮釋資料格式，出版《中文詮釋資料(Metadata) 格式彙編》，該彙編列出目前臺灣地區已設

計出來的詮釋資料格式約有 26 種，所包括的資料類型相當多元：古籍善本、古文書、書畫、文獻、拓片、古照片、雕塑、器物、戲曲唱片、皮偶劇本、客家文物、平埔文化、蝴蝶、生物、人文與自然資源地圖、國土資訊系統、權威檔（人名、主題、地名、時代）、參考書目、展覽等。（註 28）

由於台灣地區「圖書館法」在 2001 年 1 月 17 日公佈實施，其中第六條條文明訂：「圖書資訊分類、編目、建檔及檢索等技術規範，由中央主管機關指定國家圖書館、專業法人或團體定之。」因此國家圖書館依法需負責研擬圖書館資訊相關的技術規範。考量現今的大環境，國家圖書館做了以下的分析：（註 29）

- 中國機讀編目格式於 1997 年修訂之第四版與前一版在架構上有所變更，國內圖書館自動化系統至今尚未改用之。可探究美國國會圖書館推出 MARC21 之目的、過程、結果、實際運用情形與未來趨勢，做為我國修訂新版的中國機讀編目格式之參考。
- 因應網路資源的快速發展，及協助圖書館整理各種類型文獻，各種資料之詮釋格式亦相繼制定。是以如何因應新科技、新資源對於文獻著錄與描述的衝擊，整合新興資料之詮釋格式與中國機讀編目格式之異同，是個值得研究的議題。
- 國家圖書館曾於 1983 年制定「國立中央圖書館文獻分析機讀格式」，至今並未再修訂，因此修訂此文獻分析機讀格式，以作為國內文獻分析的規範，有其必要性與急迫性。
- 數位圖書館系統所運用的技術與標準多元，至今仍處於紛然雜陳的時期，網路資訊系統的建置首重系統的相容性與資源分享，必需遵守共同的標準，再尋求個別系統的特色。因此，國家圖書館實需要訂定數位圖書館資料建檔、交換與檢索標準，其內容包括：字碼標準、資訊組織標準、資料交換語法、分散檢索標準、各種類型資料數位化檔案格式等。

基於上述的背景，國家圖書館於 2001 年 4 月提出了「研擬圖書資訊相

關技術規範」計畫，分別成立修訂「中國機讀編目格式」、「文獻分析機讀格式」、「詮釋資料格式標準」、「資料數位化標準」、「分散檢索標準（XML、XSL）」及「資訊檢索服務與協定標準」等六個研訂小組，各組聘請臺灣地區相關之學者專家十至十五人為委員，定期邀請其來參加研討草案之修訂事宜。在這些小組當中，「詮釋資料格式標準」、「分散檢索標準（XML、XSL）」及「資訊檢索服務與協定標準」等小組與數位典藏的資源組織有極密切的關係。

由於 ROSS 計畫已於 2001 年春天結束，該計畫的主要成員均已參與在國家圖書館的「研擬圖書資訊相關技術規範」計畫中，將過去 ROSS 計畫的研究成果帶進此新計畫中，並以做為繼續發展的基礎。

五、結語

自從資訊科技普遍應用以來，對社會的影響日益深遠，特別是近年來電腦與通訊技術日益更新，改變了社會上各種溝通的行為，也改變了基本的資訊表達、呈現和處理的方式，其衝擊將影響到社會中的每一個人，包括的層面廣及工作方式、家居生活、教育子女和休閒娛樂的方式。因此，如何讓數位資訊妥善地被運用，在網路上開發有意義、高品質並具文化藝術性、科學教育性的內涵教材是相當重要的。

美國總統資訊科技顧問團（President's Information Technology Advisory Committee, 簡稱 PITAC）於 2001 年 2 月 9 日呈交給布希總統的“Digital Libraries: Universal Access to Human Knowledge”報告書中也強調，若能實現數位圖書館的潛能，必能使美國的各級學校擁有高品質的教育資源，因此聯邦政府應居於領導地位，並持續地提供資源以支持相關計畫的研究與發展。當圖書館、博物館、檔案館、和其他典藏單位將藏品數位化之後，他們多面臨數位館藏維護時的技術和操作面的問題。因此，為能盡速地達到取得全球人類知識的目標，PITAC 更提出了建立大規模的數位圖書館試驗環境，並支持進一步的數位圖書館研究計畫，包括：詮釋資料和其使用、規模性、互通性、數位檔案儲存與典藏、智慧財產權、隱私權和安全性、以及使用研究等

建議。(註 30)

台灣地區擁有中華文化以及台灣本土的珍貴文獻與歷史文物，是世界人類共同的無價寶藏。我們很慶幸國科會與文建會等機構在這幾年願意支持數位化典藏的計畫，提供經費進行數位博物館、國家數位典藏、以及全國文化資料庫等相關計畫。然而，一個好的數位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，它意指一個能夠整合資源、服務與人等各因素的環境，以支援資料、資訊、知識的創造、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。(註 31)

要達到這個理想的境界，就如 PITAC 的建議，支持進一步的數位圖書館研究計畫是有其必要性。由於數位圖書館最基本的條件就是需要提供方便且開放的資訊服務者給所有使用者，因此具備有良好的資源組織方式以供方便的檢索是極其重要的。台灣地區目前在進行資源組織相關的研究主要以 ROSS 計畫、中央研究院、以及國家圖書館為主。由於網路資訊系統的建置首重系統的相容性與資源分享，必需遵守共同的標準，再尋求個別系統的特色。因此，各單位如何合作共同發展資源組織與分散檢索的標準就成為當務之急。甚至如果能夠結合海峽兩岸三地的數位資源組織相關學者專家，從資訊加值的角度，共同探討處理資訊之屬性，並建立中文資訊詮釋格式，以及從資源分享的角度，共同建立詮釋資料交換之共通架構就更理想了。

註釋：

註 1：Find／How Many／上網人口／2001 年 6 月底止臺灣上網人口。

(<http://www.find.org.tw/0105/howmany/20010802.asp>) 上網日期：

2001.08.03.

註 2：Find／How Many／連網主機／2001 年春季網路調查報告。

(http://www.find.org.tw/0105/howmany/howmany_2001spring.asp) 上網日

期：2001.08.03.

註 3：陳雪華，「公共圖書館與數位圖書館/博物館的發展」，公共圖書館發展實務研討會論文集（臺中市：國立台中圖書館主辦，民國 89 年 4 月），頁 3-63—3-74。

註 4：黃鎮台，「以『人文關懷』為主軸的跨世紀科技發展」，科學發展月刊 27(7)(民國 88 年 7 月)，頁 715-718。

註 5：王美玉，「數位博物館專案計畫成果」，科學發展月刊 28(4)(民國 89 年 4 月)，頁 249-253。

註 6：陳雪華，「台灣地區數位圖書館與博物館的發展」，海峽兩岸第五屆圖書資訊學學術研討會論文集（B 輯）（台北市：中華圖書資訊學教育學會，2000 年 8 月），頁 199-200。

註 7：同前註，頁 200-201。

註 8：張善政，「數位博物館（Digital Museum）背景與構想簡報」，國科會「數位博物館」記者會簡報資料（台北市：國科會企劃考核處，民國 88 年 2 月 9 日）。

註 9：國科會數位博物館推廣辦公室，數位博物館推廣教育簡訊 15 期（民國 90 年 5 月 4 日）。

註 10：中央研究院，「國家數位典藏」國家型計畫總規劃報告（草案）（台北市：中央研究院，民國 90 年 6 月 28 日）。

- 註 11：文建會，「全國文化資料庫」計畫書（台北市：文建會，民國 90 年）。
- 註 12：陳雪華，「台灣地區數位圖書館與博物館的發展」，海峽兩岸第五屆圖書資訊學學術研討會論文集（B 輯）（台北市：中華圖書資訊學教育學會，2000 年 8 月），頁 197-207。
- 註 13：同註 12。
- 註 14：同註 12。
- 註 15：同註 12。
- 註 16：同註 12。
- 註 17：陳雪華、陳昭珍、陳光華，「數位圖書館／博物館中詮釋資料之理論與實作」，圖書館學刊 13 期（民國 87 年 12 月），頁 37-60。
- 註 18：同註 18。
- 註 19：陳昭珍、陳雪華、陳光華，「XML／Metadata 管理系統—Metalogy 之設計」，中國圖書館學會會報 65 期（民國 89 年 12 月），頁 1-12。
- 註 20：陳雪華、陳昭珍、陳光華，數位圖書館 XML/Metadata 管理系統（台北市：文華，民國 90 年）。
- 註 21：同註 19。
- 註 22：陳亞寧等著，「數位圖書博物館的資訊基礎建設：中央研究院 Metadata 的規劃設計與實施」，圖書館與資訊科學 25(2)（民國 88 年 10 月），頁 72。
- 註 23：中央研究院資訊所，「計畫簡介：數位典藏系統技術研發」，URL：<http://mars.csie.ntu.edu.tw/~dlm/plan/1st/intro6.htm>，上網日期：民國 89 年 11 月 5 日。
- 註 24：陳亞寧，「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」，中央研究院計算中心通訊 15(8)（民國 88 年 4 月），頁 70-73。

註 25：同註 24。

註 26：同註 24。

註 27：陳雪華，「電子資源組織與 Metadata 的發展」，中華民國八十九年圖書館年鑑（臺北市：國家圖書館，民國 89 年），頁 83-86。

註 28：國家圖書館 Metadata 研究小組編撰，中文詮釋資料(Metadata) 格式彙編（台北市：國家圖書館，民國 89 年）。

註 29：國家圖書館，研擬圖書資訊相關技術規範計畫書（台北市：國家圖書館，民國 90 年）。

註 30：President's Information Technology Advisory Committee, Panel on Digital Libraries. Digital Libraries: Universal Access to Human Knowledge (Arlington, VA: National Coordination Office for Information Technology Research and Development, 2001), 5-10.

註 31：Maria Zemankova, Workshop Report: Distributed Knowledge Work Environments (Santa Fe, New Mexico: Digital Libraries, March 9-11, 1997).
DIGLIB@INFOSERV.NLC-BNC.CA

參、問題研討與分析

一、數位圖書館/博物館

數位圖書館/博物館研究是近幾年才興起的研究領域，這個領域所要研究的課題到底是什麼？錢祖怡認為，數位圖書館的研究問題包括：(註1)

1. 各類型資料及各種metadata的抓取、建立、處理
2. 各種種類、格式、且大量資料的分類、組織及整合
3. 設計各種軟體及計算法則以利資料的查尋及處理，如瀏覽、查尋、過濾、摘要化及概要化等處理。
4. 發展各種工具及協定。

IITA 數位圖書館座談會則認為，數位圖書館要研究的課題及其優先順序主要有五類，其中前面三個是核心問題，也是最急迫的問題：(註2)

1. 互通性 (Interoperability)；
2. 物件的描述與儲存 (Description of Objects and Repositories)；
3. 館藏管理及組織 (Collection Management and Organization)；
4. 使用者介面及人機互動問題 (User Interfaces and Human-Computer Interaction)；
5. 經濟、社會及法律問題 (Economic, Social, and Legal Issues)。

US-EU 研究小組對於數位圖書館/博物館的研究課題也有相當深入的探討，並將研究範疇分為智慧財產權與經濟 (Intellectual Property and Economics)、全球性的資源查尋 (Global Resource Discovery)、互通性 (Interoperability)、詮釋資料 (Metadata)、多語言資訊存取 (Multilingual Information Access) 等五個部份，這五個研究範疇的關係可以用下圖表示之：

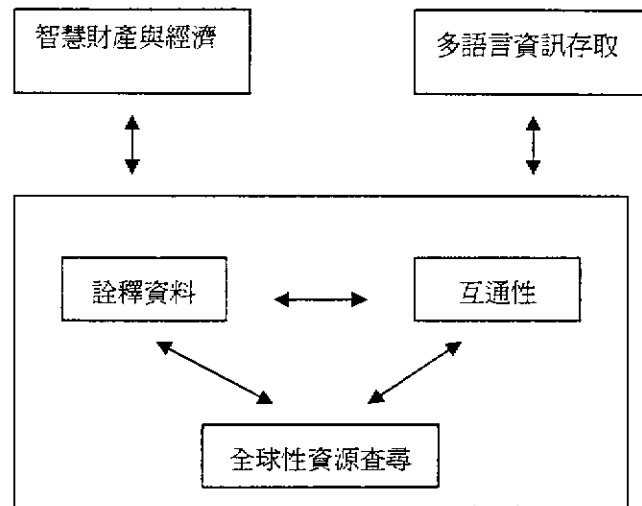


圖 3：US-EU 研究小組對於數位圖書館/博物館的研究範疇

茲將這五個研究領域的內容分別說明如下：

1. 智慧財產權與經濟(Intellectual Property and Economics)：這方面的研究課題主要包括：
 - (1) 機構與社會的政策(Institutional and social policy context)：如整個社會對電子資訊之著作權的解釋、隱私權政策等；
 - (2) 架構與機制(Architecture and mechanism)：包括如何建立相關電腦基礎建設以支援資訊經濟活動，以及如何應用相關的商業機制與政策等；
 - (3) 內容與服務(Content and services)：那些資訊是有價值的、它們如何被使用、價格如何、及其他相關問題。
2. 全球性的資源找尋(Global Resource Discovery)

在網際網路上有愈來愈大量、愈多元、愈分散的資訊，而為取用這些資訊所發展出來的協定及為何要發展這些協定的原因也愈來愈多，不僅在分享資源而已。同時，資訊使用者的數量大幅成長，背景也很多樣，從一般性的生活需求到專業性的需求都有。當網際網路被使用的越來越多元

時，資訊的品質及可靠性會越來越重要，不過資源的多樣性，也將增加此問題的複雜度。

資源蒐尋的目標不只在提高關鍵字查尋的品質，更要提高查尋的精確度與回收率，唯有高效率且高效能的分散式查尋，才可能創造出新型態的資訊利用模式。這方面的研究課題主要包括：

(1) 系統問題：包括如何發展一個可提昇查尋及導覽的系統基礎建設，以支援中介站提供線上資訊服務。此基礎建設應可辨識、查尋、檢索可用的電子資源。

(2) 內容問題：在內容方面要解決的問題包括，

a. 資料庫的選擇；

b. 多媒體資料的表達(representation)及查尋問句；

c. 語意的異質問題：單一的資料庫可能包含不同的文獻類型，不同的資料庫其資料庫架構也會不同並使用不同的問句語言，從使用者的角度來說，這些不同與他的資訊需求無關，不應該是他要面對的問題，他只要能取得合適的資源即可。所以要有一個可解決不同的資料庫語意異質的方法。

d. 排序問題：為便利使用者的利用，資料庫及文獻的排序對電子圖書館而言是必要的功能，排序的條件包括：資訊品質、合適性等，並可根據特定對象，如兒童的需求做資訊過濾。

(3) 人機介面：人機互動的問題可分為下列四個面向來看，前二者是有關輸入、輸出的問題，後二項乃是屬於認知心理的問題，其中一個在瞭解使用者想做的事，另一個在幫助使用者認識系統的處理過程及實際所事，這四個問題如下：

a. 查尋問句的建構與指引

b. 查尋結果的視覺化及呈現

c. 使用者問題的瞭解

d. 過程展開及使用者教育

(4) 組織方面的問題：在分散式查尋領域，每一個論題都會牽涉到組織層面。如建立館藏及技術支援策略等，都和組織環境有關。每一個組織都有不同的目標及優先順序，所以如何讓區域網路中的不同機構得以合作，以得到最大的效益，是一個很大的挑戰。

3. 互通性(Interoperability)

互通性對數位圖書館而言是很重要的課題，在數位圖書館/博物館的環境中，不同的系統，可以有不同的功能、不同的重點，但對使用者而言，卻希望這些不同的資源可以整合在一起使用，所以說，互通性研究是建立大型數位圖書館/博物館的基礎。所謂的互通性，牽涉到好幾個層次的問題，在作業平台層次，互通的問題包括：電腦硬體、作業系統、網路協定、網路架構等，而在資料庫層次，異質的問題包括：資料模式的差異、限制(constraints)、查詢語言、交易控制機制(transaction control mechanisms)等，這些都屬於比較低階的問題。即使這些問題都解決了，資料庫的互通還是有問題，因為還有語意異質(semantic heterogeneity)的問題，語意異質包括：屬性命名、格式、結構、流失及衝突值的確認等問題，這些都是數位圖書館/博物館中重要的研究課題。

4. 詮釋資料(Metadata)

在網路時代，詮釋資料(metadata)一詞乃指有關資料的結構性資料(structured data about data)，傳統的圖書館目錄、網頁上的書目標題等都是metadata。不同的使用者群體，如圖書館、電腦科學家、政府機構、博物館、出版社、商業機構、法律社群等，各有其不同範圍與目的的詮釋資料。在這些不同的群體當中，有些群體剛開始使用網路化資訊，也剛開始在設計其專門領域所需的metadata系統。Metadata是資源找尋、文獻傳遞、權限管理、資料存檔的基礎，對電子商務及電子出版而言，metadata也是很基本的要項。

5. 多語言資訊存取(Multilingual Information Access)

這個世界透過網際網路，已越來越緊密的連結在一起，而 WWW、電子郵件、遠距合作、數位圖書館/博物館、電子商務，以及其他類似的發展，使得各種資訊都可透過網路供全球使用。所以，在這種情況下，資訊環境的建立勢必要考慮多語文、跨語文的問題。有關這方面的研究所牽涉的問題有：

- a. 資訊檢索問題：包括自由詞索引、查詢處理、檢索機制、詮釋資料的建立
- b. 機器翻譯問題：包括語言轉換等技術。
- c. 計算語言學：包括語法分析、剖析器、詞彙辨識與擴充、語意分析、語言建立等
- d. 文獻處理：包括文獻聚集、分類、過濾等
- e. 資源發展：包括字典、索引典、索引詞、術語庫等的建立
- f. 人機互動與展現：包括文獻集的視覺化、單一文獻或多文獻的概要化等
- g. 語音處理：包括語音分析、辨識及產生等。

二、數位圖書館/博物館的資訊組織

由上述學者專家的見解中，我們可以知道資訊組織在數位圖書館/博物館研究領域的重要性。而在多元的數位圖書館及博物館世界裡，由於不同領域所要處理的資料性質不同，不可能統一採用一種 metadata 格式，不同的學科領域會採用不同的 metadata，如目前歐美國家已發展之 metadata 格式，有圖書館界的 MARC，FGDC metadata，GILS metadata，EAD metadata，CIMI metadata，Dublin Core metadata 等，共二十幾種之多。但是，雖然不可能統一採用一種 metadata，但也不應該是一個資料庫一種 metadata，因為，不同的資料庫，有匯集在一集中式目錄中查尋的需求，或是做分散式

查尋的需求。在網路環境中，分散式查尋已是必然的趨勢。為了使電子圖書館/博物館可做分散式檢索，因此我們制訂了 Z39.50 Profile，同時也將嘗試發展可以跨不同的 metadata 查尋的 Z39.50 client 與 server 端系統。

三、XML/Metadata 輸入、交換及管理系統的設計

由於 metadata 格式相當多，不可能只採用一種 metadata 格式，不同的學科領域會採用不同的 metadata，如目前歐美國家已發展之 metadata 格式，有 FGDC metadata, GILS metadata, EAD metadata, CIMI metadata, Dublin Core metadata 等共二十幾種之多。但這些 metadata 未來共通的交換語法，應該是 W3C 積極推動的 XML，因此開發一個以 XML 為核心之 metadata authoring tool，供各個數位圖書館/博物館系統使用非常重要。這個系統需要能接受各種不同的 DTD (Document Type Definition)，換言之，這個系統可以支援各種 metadata 格式，才能適用於各種主題及各種資料類型的數位博物館。

此外，我們要發展的這個系統將不僅是一個 XML authoring tool 而已，當系統轉入了一個 XML DTD 以後，它會自動轉成內部的資料庫結構，並產生輸入介面，所以事實上這是一個相當有彈性且功能完整的數位圖書館/博物館系統。此系統也可以轉入及轉出符合某種 DTD 結構之 XML 記錄，換言之，系統有 XML parser。整個系統架構如下圖所示：

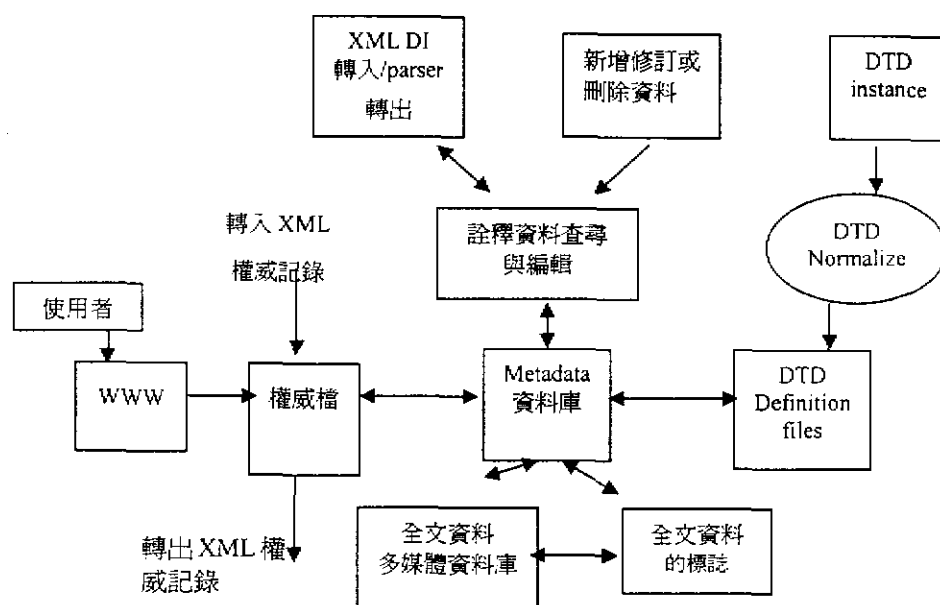


圖 4：XML/Metadata 系統架構

四、數位博物館的權威控制

實體圖書館的資料經過仔細的組織與整理，使用者通常可以很容易地取得所需要的資訊，其中關鍵的因素除了適當的詮釋資料格式之外，還有賴權威控制的概念。經過權威控制的資料，使用者無論是使用孫文、孫中山、中山樵、孫逸仙都可以檢索有關孫中山先生的資料；同理，使用寮國或是老撾也可以取得寮國的相關資料。除此之外，透過廣義詞、狹義詞、相關詞、同義詞等相關詞彙的使用，使用者也更加容易檢索資訊。為了有效達到上述的功能，必須從資料著錄與資料檢索兩個角度觀察權威控制，因此，權威控制也可視為是資料著錄與資料檢索的橋樑，下圖描述三者之間的關係。唯有在資料著錄的階段，仔細處理著錄的內容，使用者才能夠在檢索時得到上述的服務。同時，權威控制也讓資料著錄者更有效率地處理資料，降低資料處理的成本，提升資料著錄的一致性。

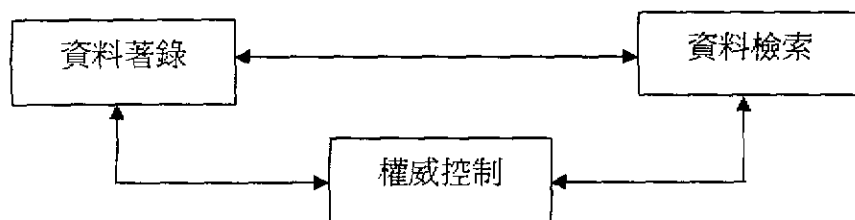


圖 5：數位博物館的權威控制

五、數位博物館的整合查詢

網際網路的出現，使得數位化資訊的使用與交換不再僅限於文字型資料，而是包括聲音、影像、或視訊等多媒體資訊；網際網路無國界的特性，也使得資訊的查尋不再僅限於某地區、某主題、或某文化，而是多媒體、多文化、多語文(Multi-Media, Multi-Cultural, Multi-Lingual)的查詢。就系統的角度來說，需要做分散式的查詢。不過若從使用者的角度而言，就是要讓他能用同一個介面做整合性的查詢。

無論從電子資源的多元性、分散性，或各領域各自發展不同的資訊組織模式，以及電子圖書館整合服務的理念而言，提供整合性的查尋介面，對使用者而言是非常重要的功能。而整合性的查尋模式一般而言，仍可分成建立集中式的聯合目錄，以及建立虛擬聯合目錄兩種。虛擬聯合目錄又可稱為分散式聯合目錄，換言之就是不建立任何實體的聯合目錄，而是透過查尋引擎做即時分散式的查尋，Z39.50 是建立虛擬聯合目錄的標準之一，也是目前在數位圖書館/博物館領域中，廣受採用的標準。

我們希望能將過去所發展出來的 Z39.50 client /server 程式做一些修改，將它與 Web 結合，並使它能適用於 MICI-DC 語意之資料庫，且能與國外的數位博物館系統一請整合檢索。這個部份的系統架構如下圖所示：

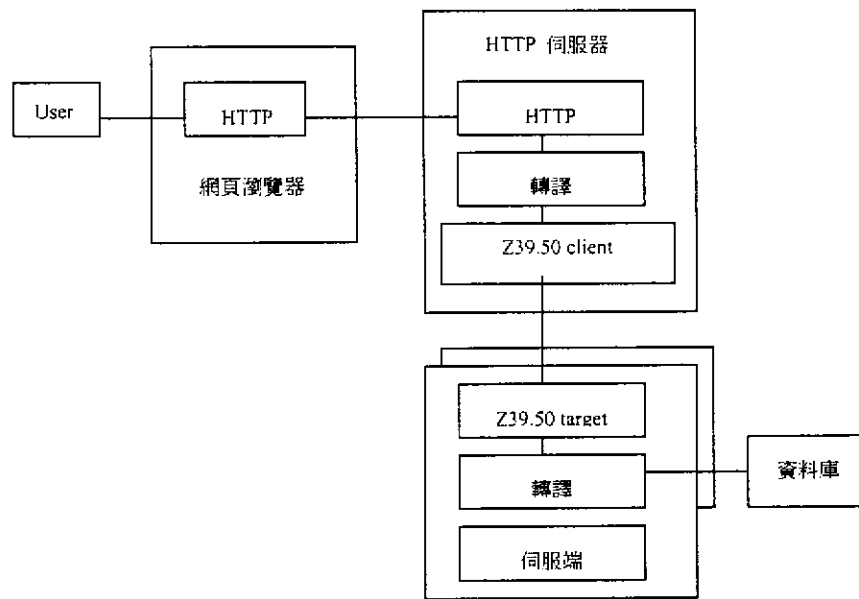


圖 6：HTTP/Z39.50 gateway 之系統架構

註釋：

註 1：錢怡祖，”展望 digital library 之未來—全球技術發展趨勢及經濟定位剖析”

Digital Library 技術發展與應用趨勢研討會 八十六年三月二十七日，經濟部技術處，資策會系統軟體實驗室主辦，[9]

註 2：Peter Schauble and Alan F. Smeaton (ed.) “An International Research Agenda for Digital Libraries: Summary Report of the Series of Joint NSF-EU Working Groups on Future Directions for Digital Libraries Research, “ (Oct. 12, 1998)
<http://www.iei.pi.cni.it/DELOS/REPORTS/Brussrep.htm>

肆、中文索引典的研究與應用

一、數位圖書館與數位博物館

數位圖書館／博物館的環境中，有效地提供資訊的利用，良好組織與整理的工作是極必要的。圖書資訊學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、標題表以及索引摘要等方式，因此圖書資訊學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，除了圖書資訊學界人士之外，還有電腦與資訊工程學界或是各學科領域的專家們，紛紛探討此相關議題，並制定各種詮釋資料（metadata）以因應資訊檢索之需。這標誌著長期以來圖書資訊學界的作法，已經獲得資訊工程學界的認同，亦即「適度地控制資訊」是提供高品質資訊的有效作法。同時，許多學術會議也將重心放在相關的議題，例如 1999 年度計算語言學會議的一項重要議題即為「控制詞彙理論與應用」；IEEE（Institute of Electrical and Electronic Engineers）與 ACM（Association for Computing Machinery）等資訊工程學界重要的學會亦舉辦詮釋資料學術會議，推動資訊整理與組織的研究，期望能夠適度地控制資訊。（註 1）

基本上，詮釋資料是敘述資源屬性的資料，通常稱為“Data about data”或是“Data describes other data”（註 2）、“Additional information that is necessary for data to be useful”（註 3）、“information about the data that helps in optimization and management of that data”（註 4）等廣狹不一的定義。吾人可以說詮釋資料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。（註 5）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詮釋資料，也是不可或缺的工具。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。由於詮釋資料是經由人工（館員或專家）判讀，將藏品中的重要資訊抽離或標示出來，這個動作賦予詮釋資料之語意，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。數位圖書館／博物館蒐藏的資源，一般而言皆為比較珍貴的藏品，因此花費昂貴的人力進行資源特性的詮釋自然是必要的。由此可知詮釋資料在數位圖書館／博物館中占有相當重要的地位。全文檢索的缺點在於沒有進行權威控制（Authority Control），因此可能會發生檢索不到或是檢索到非所需的資訊，因此詮釋資料可與全文檢索相輔相成。

由於資源的詮釋資料構成了數位圖書館／博物館系統重要的一項資源，如何檢索這些詮釋資料，進而檢索符合所需的數位典藏，就是數位圖書館／博物館系統研究的重要課題。一般而言，數位圖書館／博物館系統用以描述珍貴典藏的詮釋資料，有所謂的權威控制的觀念，如人名的權威控制、地名的權威控制、主題概念的權威控制等等。本研究著重於主題概念的權威控制，企圖補足以往資訊檢索系統無法顧及的權威控制的功能，這卻是數位圖書館／博物館系統極為重要的一環。本研究建構的資訊檢索系統的服務子系統，也就是提供一般的全文檢索系統，能夠將使用者下達的檢索詞彙轉換為系統使用的控制詞彙，而這些控制詞彙即是系統用以描述藏品主題概念的詞彙。索引典與標題表就是這些控制詞彙構成的集合，因而本章將以文獻分析的角度探討索引典的功能，及其在資訊檢索系統扮演的角色。

二、索引典

索引典為控制詞彙的構成的集合，詞彙之間有語意上的相關性，其涵蓋範圍通常以特定學科領域為主。索引典明確地整合概念，以詞彙上的控制來轉化文件所使用的自然語言，所以可用以轉換使用者的檢索詞彙為系統的索引詞彙。索引典的詞彙關係包括三種類型：等同性（Equivalence）、階層性（Hierarchy）及關聯性（Association），請參見圖 2.1。索引典有以下的功能：（註 6）

1. 傳達意義、定位及結構：協助使用者思索問題、以更好的查詢模式提供幫助、需要有好的架構來展示，很多索引典都只展示部分，而不能看出整體的結構。索引典要做有意義的排列，在線上環境沒有必要依字母順序排。

在發展有意義的架構上，還需要投入更多的研究。

2. 提供詳盡的關係（舉例 P16-）和定義：索引典應該提供完整的定義，而不只是註解（notes）而已。要提供詞彙的各種意義。
3. 以知識為本，提供索引及檢索的協助。檢索：同義詞擴展（query term mapping）、階層式擴展；索引：可協助人工索引，亦可改善自動索引（例如：MedIndex）、做自然語言處理時，使用結合語言學字典與索引典的工具（例如：UMLS）
4. 將文本與索引典做連結、連結不同的索引典：協助使用者在閱讀時可透過索引典了解定義，並透過階層式的構造獲得概念。在不同的資料庫中檢索時，連結不同的索引典很有幫助。理想是：有一個虛擬的索引典能連結到各種不同的索引典、字典或其他工具，可以提供詞彙的整合性概念。
5. 協助使用者維護他們自己的索引典。索引典建構的程序通常分為詞彙的產生、詞彙的整合、詞彙的分類、詞彙關係的建立，簡述如下：（註 7）

(1) 詞彙的產生

由相關的資源（如字典、辭典、書後索引或主題詞表）蒐集詞彙，或由範圍更廣的索引典蒐集部分的詞彙做為基礎以擴展專門學科的索引典，或由學科專家制定詞彙。

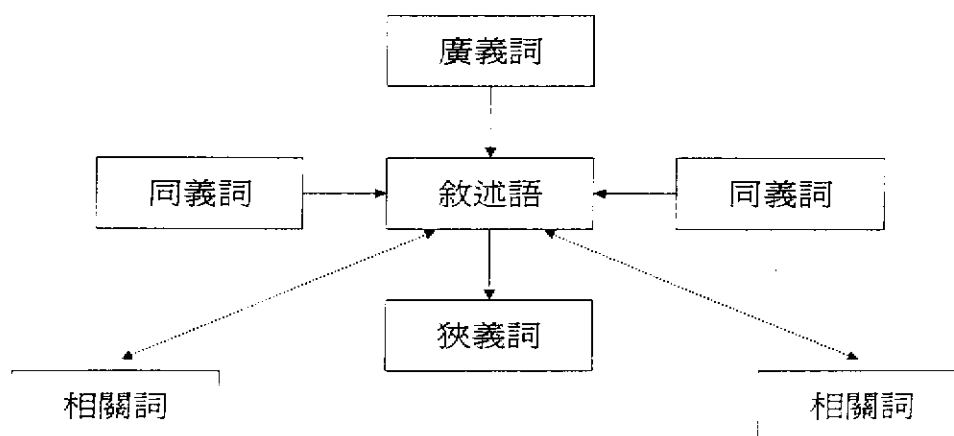


圖 7：詞彙的語義關係

(2) 詞彙的整合

詞彙產生後必須經過整理與合併的程序。所謂的合併是消除重複出現的詞彙，接著處理同義詞與近同義詞，並挑選各概念的代表詞彙。因而產生二種詞彙：敘述語（Descriptor）及款目詞（Entry Term），而敘述語即為權威控制詞彙，款目詞則為引導至適當敘述語的詞彙。

(3) 詞彙的分類

在代表主題領域的主要敘述語都決定之後，這些詞彙皆必須依其內容意義再做邏輯的分類，其目的在於將同一類之詞彙連結以方便檢索，同樣的，不同類別中的詞彙彼此有所關聯者亦應建立連結。分類有許多方式，可向學科專家諮詢，或集合各領域的專家共同協商討論，雖然專家的意見都較為主觀，但此方法尚能反應該領域使用者的資訊需求。另一方法為蒐集該學科領域的文件，以統計的方式分析詞彙的關聯。

(4) 詞彙關係的建立

索引典中有兩種不同的語意關係，一為階層關係，亦即建構詞彙間的廣狹義關係。一為並行關係，並行關係有二種型式：等同關係與關聯關係。凡具有等同關係的詞彙，彼此在概念上是相同或是在用法上是相同的，但關聯關係的建立常受主觀影響。

索引典建構了詞彙的關係，資訊檢索系統如何有效地使用索引典，或是更進一步整合索引典於系統之中，是一項重要的研究課題。傳統的資訊檢索的研究或是檢索系統的開發並不重視索引典的功能，在進入數位圖書館／博物館的研究領域後，這個觀點仍然未被廣泛接受，然而由於數位圖書館／博物館典藏品的特性，加速索引典與資訊檢索系統的整合益形迫切。

資訊檢索系統的目的是希望使用者能藉以檢索所需的資訊，過濾無用的資訊。因此系統中的資訊必定是經過選擇、徵集與概念分析的過程後，將其加以組織整理並儲存，令使用者能正確迅速地取得所需資訊。目前使用者與各種資訊檢索系統之間的溝通模式與互動，以文字詞彙為主，除了利用關鍵詞進行全文檢索

外，有些資訊檢索系統在資訊組織與整理的過程，針對文件的內容進行分析，給予文件資料檢索標識(主題詞彙或分類號)，並使用轉換後的索引詞彙。相對地，資訊使用者提出查詢問句後，也經過同樣的概念分析與詞彙轉換的過程，再與資訊檢索系統互動以找尋所需資訊。亦即，資訊使用者與資訊檢索系統之間，是藉由詞彙間的對映來達到擷取與過濾資訊的目的。

資訊檢索系統中，文件與查詢問句所具有的相同概念可經由詞彙顯現，且詞彙本身的語意關係(如相關詞、狹義詞及廣義詞)亦於概念間建立連結。因此，詞彙不僅表現出文件的內容，同時也表達檢索者的查詢問句。一般而言，如果檢索者可以將問題以正確且適當的詞彙表達出來，而系統也可以將之轉換為相同概念的詞彙，則檢索結果應能夠滿足使用者的資訊需求。因此，理想的檢索過程應該是以極為準確的方式處理「提出問題→選擇檢索詞彙→進行檢索→提供答案」的程序。(註 8)然而，實際上目前資訊檢索系統所呈現的檢索結果並不理想，其原因為使用者無法確切了解自己的資訊需求，且通常缺乏足夠的知識與能力找出適當的詞彙表達查詢問句，再加上資訊系統對使用者的資訊需求不甚了解，以致使用者所下達的檢索詞彙常與系統使用之索引詞彙不同，而造成檢索效益無法提升。

通常使用者在進行資訊檢索時，並不是那麼確切地知道想要找的文獻篇名是什麼、作者是誰，只知道主要是哪一類的文獻，在對自己的資訊需求不夠了解的情況之下，很難得到令人滿意的結果。此時，檢索系統可以透過引導的方式，在數個跟搜尋主題相關的層面中顯示線上索引典，讓使用者自行做選擇，以幫助使用者把搜尋主題變得更明確，如此一來便能得到較符合使用者資訊需求的檢索結果(註 9)。以「教育」為例，當使用者在檢索主題中輸入“Education”時，系統顯示進一步做選擇的層面(圖 2)。

Subject field of search: **Education**
Indicate limiting aspects to be used:

- Level
- Ethnic origin of students
- Giftedness/handicap of students
- Subject
- Country
- Public/private

圖 8：“Education” 主題之下可進一步選擇的層面

若使用者選擇“Level”，系統則自動從索引典中找出跟“Level”相關的敘述語，其他選項亦同，有的還能顯示出詞彙間的層級關係，如圖 2.3 所示，“Chinese American”和“Japanese American”皆為“Asian American”的狹義詞。

Level

- Preschool
- Kindergarten
- Elementary
- Secondary
- Higher

Ethnic origin

- African American
- Asian American
 - Chinese American
 - Japanese American
- ...
- Gypsy

圖 9：供使用者選擇的敘述語

經過使用者與系統數次的互動之後，檢索條件最後呈現在系統上（圖 2.4），由於都是從索引典中選擇出來的詞彙，在這個過程中便自然地幫助使用者將腦中的想法轉換成控制語言了。

Subject field of search: Education	
Indicate limiting aspects to be used:	
▪ Level	Elementary
▪ Ethnic origin of students	African American
▪ Giftedness/handicap of students	
▪ Subject	Reading
▪ Country	U.S.
▪ Public/private	

圖 10：經使用者與系統互動後的檢索條件

然而，使用者很難建構出適當的查詢問句，因為對該學科領域的認知或許不深，就算是該學科領域的學科專家，也可能不願意花時間將每個詞彙全部鍵入系統，總之，使用者於檢索資訊時，通常只提供大約可敘述其需要的詞彙即停止，幾乎都是很短的查詢問句，故得到的檢索結果也較差。因此，實有迫切的需要於系統上加強協助使用者建構查詢問句的功能，至於如何加強使用者與系統間的詞彙溝通並豐富使用者的查詢問句也成為資訊檢索專家努力的目標。

加拿大的 Chander 等人設計了一個以詮釋資料為基礎的圖書館專家系統，具有編目和檢索 WWW 上數位圖書館文獻的功能。(註 10) 此專家系統透過主題索引典並模仿參考館員進行參考諮詢的模式來提供服務。在編目方面，當編目人員鍵入詞彙時，系統會自動提供索引典中的控制詞彙供編目人員參考，同樣地，此專家系統能引導使用者從線上索引典中選擇最適當的詞彙來進行檢索，這樣就可以保證所使用的都是控制語言。目前這個系統只應用在資訊科學以及電子工程方面而已。

線上索引典詞彙展示的方式目前主要仍是按照字母順序排列，有主題分類架構的索引典，在各類別之下也是按字母順序排列。然而，有些學者認為理想的索引典顯示格式應該做有意義的排列，在線上環境並沒有必要依字母順序來排。(註 11) 黃慕萱教授更提出讀者導向趨勢的索引典，她認為索引典應放棄追尋標準或典型的使用者，最好能模擬每一位不同使用者的思路，讓每一位使用者都能感受到這套索引典是專門為他所設計的。(註 12) 目前資訊科技雖然一日千

里，但在發展有意義的索引典架構上仍需投入更多的研究。

由於使用人工分類成本太高（一篇期刊文章要\$1.7，1 億 7 千 8 百萬個網頁就需要將近三億元的分類費用），因此 Northern Light 設計了一個人工智慧系統，使用控制詞彙來做網頁的分類。分類架構及詞彙參考 DDC, LC Subject Headings, UMI Subject Headings 等。（註 13）其做法如下所示：

1. Mixed (vs totally manual, totally automatic): human-directed
2. Based on words contained in document
3. Uses Term Frequency/Inverse Document Frequency methods to match document to term(s) from controlled vocabulary
4. Each term has set of co-occurring terms derived from training set
5. Document must have a strong degree of 'aboutness' to class

近幾年的研究趨勢顯示：索引典的研究與索引典與資訊檢索系統逐漸受到研究者的重視。因此充分運用索引典的功能，加強資訊檢索的品質，遂成為一項研究課題，其中最重要的便是查詢問句的擴展與檢索詞彙／索引詞彙的轉換。

三、查詢問句擴展

資訊檢索的目標即是擷取出最符合使用者資訊需求的文件。乍看之下，達到此目標似乎很容易：只要文件中包含使用者所提供的檢索詞彙即被判斷為相關文件。然而，一般而言，網際網路搜尋引擎的使用者所提出的查詢問句平均都不超出二個詞彙，檢索詞彙的不足實難用以判斷文件的相關性，而且似乎有許多相關的文件則因為未包含與查詢問句相同的詞彙而遭忽略。一般而言，檢索效益皆隨著查詢問句的增大而提升，因為較長的查詢問句提供較多的資訊予系統進行判斷與檢索；相反的，較短的查詢問句沒有足夠的資訊，則檢索的結果即不佳。

以向量空間模型為基礎的資訊檢系統是利用計算查詢問句與文件中對應詞彙的權重來評估其相關性，但如果文件本身相關，而其所使用的索引詞彙不是與檢索詞彙相同而為檢索詞彙的同義詞，則會有判斷錯誤的情形。一個可能的解決

辦法即是擴展查詢問句，以更多相關詞彙提供足夠的資訊。擴展的方式除可利用相關回饋之外，亦可利用索引典中詞彙的同義詞關係來進行查詢問句擴展，其中所謂同義詞可指十分相關的詞彙，亦可謂於文法上或語意上完全可相互取代的詞彙。Gauch 與 Smith 於 1993 年的研究中，以線上索引典來修飾查詢問句，對檢索結果有所提升。(註 14) 圖 5 描述查詢問句的機制。

查詢問句擴展的方法主要有三種：以查詢問句為基礎的查詢問句擴展、以語料庫為基礎的查詢問句擴展、以語言分析特性為基礎的查詢問句擴展。(註 15)

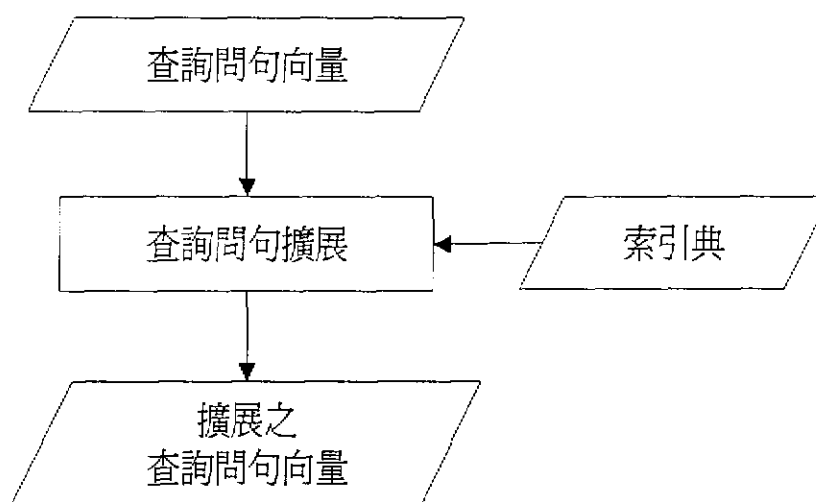


圖 11：查詢問句擴展程序

以查詢問句為基礎來擴展檢索詞彙，是從原始查詢問句的檢索結果中，擷取出文件中相關的詞彙，以進行查詢問句擴展。這種擴展的方式結合相關回饋系統，是由使用者判斷文件或詞彙的相關性。Xu 與 Bruce Croft 於 1996 年提出以查詢問句為基礎的擴展方式，可省去使用者判斷的步驟，即系統將檢索結果中前 N 篇文件做相關回饋，其中 N 為 20 到 100，檢索結果提升了 23%。(註 16)

利用語料庫為基礎來擴展查詢問句，則以分析全文資料庫中的內容為主。語料庫的分析多以詞彙共現的特性來聚集相關詞彙，或是結合詞彙共現與文法分析兩種方式，其中詞彙共現可定義為同時出現在同一篇文件中，或出現在同一篇文件的同一小節中，或是同時出現在固定且逐字移動的文字框 (Text Windows) 中，語料庫的分析則需要大量且密集的計算。

以語言分析特性擴展查詢問句，則是利用通用索引典進行查詢問句擴展，如 Longman 的當代英文詞典或是 WordNet (註 17)。然而，通用索引典中大部分的詞彙皆代表許多不同意義，系統無法判斷出最適合的詞義，難以直接應用於資訊檢索系統。

若以使用者的角度來看，查詢問句擴展技術可分為半自動 (Semi-Automatic) 與自動 (Automatic) 擴展。半自動的查詢問句擴展技術要求使用者對檢索所得的文件進行相關判斷後再回饋給系統修飾原始查詢問句，而使用者與系統的互動持續到使用者滿意其檢索結果後才停止。半自動查詢問句擴展主要的優勢為減輕使用者重新組織查詢問句的負荷，而其缺點為使用者在得到良好檢索結果之前，必須花時間與精神來判斷文件的相關度。另一種半自動查詢問句擴展的技術為結合可瀏覽的索引典。使用者先給予系統一組查詢問句，系統即提供相關詞彙清單提供使用者挑選以取代或增加原有的查詢問句，而這種方式仍需要使用者對於該學科領域的認知以及使用者與系統的互動。

自動化的查詢問句擴展技術則不需依賴使用者的相關判斷，其通常是以語言分析或詞彙共現等技術為基礎。語言分析的方法要求對查詢問句及文件有深入的了解，必須經由大量的計算來完成，然而語言分析方式的檢索效益並不顯著。(註 18)

Qiu 與 Frei 認為自動化查詢問句擴展的技術可分為四類：詞彙共現法 (Simple use of co-occurrence data)、文件分類法 (Use of document classification)、語法分析法 (Use of syntactic context) 及相關資訊回饋法 (Use of relevance information)。其中詞彙共現法的查詢問句擴展方法是依詞彙與其他詞彙間的相關程度來進行詞彙分類；而文件分類法則將文件先依文件分類演算法進行分類，其認為出現在同一類文件的詞彙彼此間有某程度的相關；另外，語法分析法方式中，詞彙關係是利用語言學知識與詞彙共現的統計分析來建立，於查詢問句擴展的應用上即於原查詢問句中加入最相關的詞彙；相關資訊回饋法則包括相關回饋技術。(註 19)

其中，索引典則為查詢問句中詞彙的重要來源，索引典可提供概念詞彙之同義詞、近同義詞、反義詞、廣義詞、狹義詞及相關詞等，但是索引典通常是由學科專家或索引專家所挑選的控制詞彙，未受訓練或不熟悉的使用者很難靈活運

用索引典。一般而言，索引典仍是資訊檢索的重要詞彙資源，具有豐富的資訊與價值，然而，目前缺乏適切的技術與方法妥善應用索引典，因此為求解決檢索上的詞彙問題，仍不能忽略索引典在查詢問句擴展上的運用。

索引典於資訊檢索系統上的應用具備兩個重要的基礎，其一為概念結構，其二為詞彙控制。所謂概念結構可分二方面來說，在索引方面，是指將文件適當地加以描述或標明；在檢索方面，是指可以從很多不同的角度及觀點上搜尋文件資料。至於詞彙控制方面，由於無論在索引或檢索方面，皆可能有多個不同詞彙可代表同一概念的情形，需要詞彙控制的機制令索引者及檢索者無論以何種詞彙切入都會被引導至適當的概念，故詞彙控制可確保索引及檢索相同概念時，會使用同一個詞彙。(註 20)

索引典的功能如上述，是為詞彙控制的資訊檢索系統提供標準化的詞彙。至於詞彙的控制則應用索引典以控制同義詞、近同義詞、分辨廣義詞、狹義詞及相關詞等，不但可據以進行文件描述，亦提供予使用者適切的檢索詞彙。索引典於檢索系統上的應用，除了可使系統索引詞彙與使用者檢索詞彙達一致性，且能提供概念詞彙之間的各種關係，建議更多更適切的檢索詞彙。然而，索引典有易於老化過時的缺點，且檢索詞彙因局限於索引典，亦有狹隘的缺失。

總之，索引典所提供的控制詞彙對於資訊檢索效益的提升具有重要的地位，且索引典亦能依據其詞彙間的關係與連結為使用者進行相關判斷，以減低使用者檢索資訊時所面臨的困擾。

英文索引典於資訊查詢問句擴展方面的研究與發展已有一段時間，也有不少的成果。Stiles 是最早提出利用相關詞彙來改進檢索效益理論的學者之一。Doyle 進一步提出無論人工或機器判斷出來的相關詞彙都應加入檢索。Courtial 及 Pomian 則強調在科學與技術領域中使用者需要的檢索詞彙及概念常超出傳統索引典中所能提供的資源。(註 21)

許多研究亦合併索引典與資訊檢索系統。如 CODER system 採用 Handbook of Artificial Intelligence 及 Collin's Dictionary。Ahlsweide 與 Evens 分析 Webster's Seventh New Collegiate Dictionary 以建構詞彙資料庫，其包含詞典及語意關係。(註 22) Lesk 將一電子版的 Murray's Oxford English Dictionary 轉換成索引典形

式的工具來輔助檢索。(註 23) 前述的種種作法皆傾向於建構通用詞典而非專門領域的索引典。Chen 及 Phar 整合 LCSH 的部分詞彙於系統，利用 Branch-and-Bound 擴展演算法協助使用者陳述查詢問句，(註 24) 且 Chen 及 Phar 更建立以概念為基礎的文件檢索系統，採用多個索引典，包括 LCSH、ACM Computing Review Classification System、自動建構的電腦科技索引典等等。(註 25)

Crouch 及 Yang 從文件關鍵詞自動產生索引典的類別，其可利用於文件的索引及查詢問句的建構。Crouch 的方式是基於 Salton 的向量空間模型及詞彙識別理論，文件是以 Complete Link Clustering 演算法來分類。(註 26)

美國國家醫學圖書館的 Unified Medical Language System (UMLS) 是目前專門領域之索引典中建構與發展最為迅速完整，系統目標在於建構一智慧型的自動化系統，可理解生物醫學的專有名詞與各詞彙間的關係，使用者可透過此系統的協助來組織、檢索機器可讀的資訊資源。UMLS 包括索引典的索引典 (Metathesaurus)、語意網路 (顯示 Metathesaurus 中類別間的關係)、資訊資源指引 (包含各生物醫學資料庫的資訊)。(註 27)

近來許多研究亦結合既有的多種索引典來輔助資訊的檢索。Chamis 即探討有關索引典的合併、系統與策略的建立來解決資料庫不相容的檢索難題，其提出的 Vocabulary Switching System (VSS) 系統整合 12 個索引典，包含商業、社會科學、生命科學與物理學四個學科領域。(註 28) Knapp 的 BRS/TERM 詞彙資料庫可對映自然語言同義詞與控制詞彙的敘述語，該索引典應用於 7 個書目資料庫，包含社會科學與行為科學領域。(註 29) NTIS 資料庫則包含多種政府機構的資料，各個資料庫都有專門索引典，NTIS 索引典則合併整合各 Subthesaurus 的詞彙，於 NTIS 索引典中的每個詞彙皆加上其原屬於何 Subthesaurus 的標記。(註 30)

綜合過去的研究，索引典應用於資訊檢索效益的提升上，大致依下列方式進行：

1. 互動式瀏覽

於資訊檢索的過程，可提供階層式的詞彙表單讓使用者逕行挑選，以重組查詢問句。CANSEARCH 為一書目檢索系統，原是为提供醫師檢索於 MEDLINE 中的醫學文件而設計，此系統包含該專門領域的知識，即採用互動式瀏覽提供索引典資源。(註 31) IR-NLI II 利用線上索引典作為系統與使用者的溝通橋樑。檢索的進行則由使用者輸入一串詞彙，系統即依據使用者的輸入提供該詞彙於索引典中的位置與鄰近的詞彙予使用者參考。(註 32)

2. 索引典結合相關回饋

I³R 採用使用者檔案與相關回饋，亦利用使用者與系統的對話來建構查詢問句，提供自然語言檢索介面，系統以索引典於查詢問句修飾程序中提供相關詞彙，亦提供初次檢索的相關文件清單，由使用者從中挑選相關文件回饋予系統進行查詢問句的重建。(註 33) Shoval 建立一系統來幫助使用者挑選適當的檢索詞彙。於語意網路中收錄了詞彙、概念與片語彼此間的語意關係，系統即利用這資源為使用者提供適切的檢索詞彙，進行查詢問句的擴展，然而此系統需要使用者對於所建議之詞彙做相關判斷，亦即使用者必須親自修飾原查詢問句。(註 34)

3. 自動對映

於通用的資訊檢索系統架構上加入自動查詢問句擴展的機制，為先給予系統查詢問句，系統摘出問句的關鍵詞後，即於索引典搜尋相關的詞彙，再自動將追加的詞彙與原查詢問句結合，形成擴展後的新查詢問句。

四、中文索引典的應用

本節將說明中文索引典的研究與應用的情況。目前部分學科領域已經發展出專門領域索引典，但多以利用索引典來描述文件，如立法院的立法資訊系統索引典、農資中心的農業科技索引典、科資中心的科技索引典等，都是根據既有國外的索引典再加上中文特有詞彙組合而成。以下就立法院、農資中心及科資中心之索引典分述如下：(註 35)

1. 立法資訊系統索引典

立法資訊系統索引典是由「立法院資訊及研究服務小組」所編訂，此索引典是將美國國會圖書館國會研究服務處(Congressional Research Service, 簡稱 CRS)所修訂的第十六版「立法索引詞彙」(Legislative Indexing Vocabulary)，又名「美國國會研究服務索引典」(The CRS Thesaurus)譯成中文，並參照美國國會研究服務處選定索引詞彙的步驟與方法，有系統地從國內各類立法文件資料中，挑選意義明確的詞彙以進一步擴充。立法資訊系統索引典為立法院資訊系統控制詞彙之索引與檢索工具，應用於立法委員質詢與答覆資訊系統、立法期刊文件資訊系統、立法院法律全文資訊系統、立法院法律沿革資訊系統、立法新聞資訊系統、立法院圖書館自動化系統及立法議事紀錄索引系統等七個資訊系統中。

2. 農業科技索引典

農業科技索引典由農資中心所編訂，初步將聯合國農糧組織資訊系統(AGRIS)之索引典 AGROVOC 譯為中文，並加入國內相關文件的主題詞彙而完成。農業科技索引典包括紙本式與電子版，其中紙本式索引典共分為三部分：(1)索引典詞彙 (2)中文索引 (3)英文索引，電子版則與農資中心的資訊服務系統結合，可進行線上查詢。

3. 科技索引典

科技索引典由國科會科資中心所編訂，其參考日本科技資訊中心的 JICST 索引典(JICST Science and Technological Thesaurus)、美國國家醫學圖書館的 MeSH、英國國家標準局的 ROOT 等國際上知名索引典之優點所建構，中英文並列。目前該索引典所涵蓋的主題共 21 類，大類之下又細分為 182 個類目。

上述索引典應用多以系統文件的描述為主，其中農資中心的系統完全採用控制詞彙，立法院與科資中心的系統則除了索引典的控制詞彙外又加入自然語言，以增加其求全率，立法院資訊系統的自由詞彙則是利用人工從書名或內文中擷

取，而科資中心的自由詞彙則是由索引者讀完文件後，自文件中摘出適當的關鍵詞。

資訊科技促使圖書館的自動化與線上系統的使用，用以組織、整理並呈現資料的目錄亦必須隨著潮流與需求進行轉變，因此有關線上公用目錄方面的研究大受重視，但線上索引典卻相對地受到忽視，目前的線上索引典相較於紙本索引典似乎只是承載媒體的不同而已，並沒有因為電子化而較能發揮索引典的價值。既有之中文索引典大多先以大型且多冊的紙本型式出版，所包含的詞彙從數千到數十萬不等，一些重要的中文索引典也已電子化，並與檢索系統相互結合，但目前並沒有顯著的成果。

事實上，文件的描述和查詢問句的陳述階段裡都會有一定程度的不確定性存在。詞彙的選擇可能是隨機的，實無法預知使用者到底會選擇那些詞彙，因此即使使用非常嚴謹的索引典，一本書還是可以用很多詞彙來描述，而且通常索引者和檢索者不見得會使用相同的詞彙來描述同一本書。

檢索過程中亦很難認識詞彙的複雜性和彼此之細微區分。通常檢索者必須經過一連串的過程才能確定該使用那些檢索詞彙，此過程包括對系統文件描述方式的認知、對索引語言的認知，以及經由詞彙的語意關係匯集更多詞彙意義等。使用者在下達查詢問句時，應盡可能將所有相關詞彙列出，以目前線上檢索系統而言，一方面系統很少提供這種功能，一方面大部分的使用者不知道應盡可能將相關詞彙列出才能提高檢索品質，且即使系統提供線上索引典以輔助查詢問句的建構，使用者通常也不知道有這樣的功能可利用。(註36)

中文環境中，索引典最初的應用方式是讓使用者自行選擇概念相近的詞彙，系統本身並不存在規則或機制以協助使用者找到適合的檢索詞彙，而是提供「候選」的詞彙讓使用者挑選。因此目前大部分的系統利用統計方式找出適合的詞彙，幫助使用者建構並改善其查詢問句，或是將紙本的索引典內容電子化，直接於查詢進行將適合的詞彙表列給使用者參考。(註37)

前述種種的情況都表露中文資訊檢索研究上太忽略索引典的價值，再看英文索引典應用的研究成果，更顯現中文索引典發展與應用之迫切需要。中文與英文於資訊檢索上最大的不同為斷詞的處理，英文詞與詞之間有空格區隔，而中文詞

彙之間並沒有特別的標示，因此即產生了斷詞的問題。除此之外，中文資訊檢索亦因語言的特性遭遇更多困難，如同義詞過多、中文碼不統一、繁簡異體字的使用、個人用詞習慣與文法上的差異、翻譯詞彙不一致等，所以更需要索引典的輔助以整合詞彙資源與詞彙間的關聯，才能發展出高效率的中文資訊檢索系統。(註 38)

中文索引典於資訊查詢問句擴展的應用，可參照 SMART 系統所採用的方式，其利用索引典將相同概念的詞彙以階層式的方式集合在一概念敘述語之下。或是將索引典視為一語意網路 (Semantic Net)，則詞彙代表某一概念，詞彙與詞彙之間的關係形成一個概念的網路，在擴展查詢問句利用這種資源，即可將檢索詞彙放入此語意網路中，所以，距離檢索詞彙最「近」的詞彙就是最相關的詞彙，這概念加以擴充後，則可在網路的連結 (Link) 上加上權重，再整合所有資訊重組原始查詢問句。(註 39)

中文索引典建構的成果或許仍不能滿足目前與未來的資訊需求，故可能需要結合兩種以上的資源來支援查詢問句擴展，結合則必須先建構詞彙間互動與連結的方式，至於可能的方式有聯合 (Union)、連結 (Chaining) 與對映 (Mapping)。聯合為查詢問句個別在不同資源進行查詢問句擴展後將兩組擴展的詞彙組合，整合所有出現於各組合中的所有詞彙，形成擴展後的查詢問句，聯合有兩種方法可結合不同資源：加權與不加權。加權指各組合皆附帶權重，兩組詞彙交錯者其權重相加。連結策略則依次使用不同的資源，即查詢問句先利用一資源後，將已進行一次擴展後的查詢問句利用另一資源再次進行擴展。對映則將查詢問句分別利用不同資源進行擴展，再擷取二組擴展後之查詢問句重疊的部分，以完成查詢問句的擴展。(註 40)

註釋：

- 註 1： ACM 與 IEEE 紛紛舉辦詮釋資料相關會議，請參見下列網站。Conference on Digital Libraries: Post-Conference Workshops, 1997,
<<http://www.sis.pitt.edu/~diglib97/Workshops.htm>> (13 Nov. 1998). First IEEE Metadata Conference, 1996,
<http://www.llnl.gov/liv_comp/metadata/events/ieee-md.4-96.html> (13 Nov. 1998).
- 註 2： Terry Kuny, "Metadata: What is it?" 24 Apr. 1997.
<DIGLIB@INFOSERV.NLC-BNC.CA>
- 註 3： Terry Kuny, <Terry.Kuny@xist.com> "Second IEEE Metadata Conference," 2 Dec. 1996. <DIGLIB@INFOSERV.NLC-BNC.CA> (2 Dec. 1996)
- 註 4： A. Perkins, "Developing a Data Warehouse"
<<http://www.ies.aust.com/~ieinfor/dw.htm>>
- 註 5： 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>>
- 註 6： Dagobert Soergel, Enriched thesauri as networked knowledge base for people and machines. URL: http://www.dtic.mil/cendi/pres_arc.html (1 Aug, 2000)
- 註 7： Miranda Lee Pao, Concept of Information Retrieval (Englewood, Colorado: Libraries Unlimited, INC., 1989), pp.122-123.
- 註 8： W. Bruce Croft, "Knowledge-Based and Statistical Approaches to Text Retrieval," IEEE Expert (1993): 8-12.
- 註 9： 同註 6。
- 註 10： Chander, P. G.; Shinghal, R.; Desai, B. C. "An Expert System to Aid Cataloging and Searching Electronic Documents on Digital Libraries."

Expert Systems With Applications 12.4 (1997): 405-416.

註 11：同註 6。

註 12：黃慕萱，「線上索引與顯示格式之設計探討」。中國圖書館學會會報 53 期（民國 83 年 12 月），頁 125-135。

註 13：Joyce Ward., Indexing and Classification at Northern Light. URL:
http://www.dtic.mil/cendi/pres_arc.html (1 Aug, 2000)

註 14：Gauch, Susan and John B. Smith, "An Expert System for Automatic Query Reformation." Journal of the American Society for Information Science 44, no.3 (1993) : 133.

註 15：Susan Gauch and Jianying Wang, "Tuning a Corpus Analysis Approach for Automatic Query Expansion," <<ftp://ftp.tisl.ukans.edu/~sgauch/papers/ITTC-FY97-TR11100-2.ps>> (20 Oct. 1998).

註 16：J. Xu and W.B. Croft, "Query Expansion Using Local and Global Document Analysis." In Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Zurich, Switzerland, Aug 18-22, 1996, 10.

註 17：WordNet 為計算語言學領域重要的研究資源，WordNet 將英文的名詞、動詞、形容詞及副詞有系統地組成同義詞組，每一詞組代表一概念，詞組之間有各種關係的連結。<<http://mind.princeton.edu/wordnet/>> Cognitive Science Laboratory Princeton University.

註 18：同註 16。

註 19：Yonggang Qiu and H.P. Frei, "Concept Based Query Expansion." In Proceedings of the 16th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval Pittsburgh, PA, USA, Jun 27-Jul 1 1993, 160-161.

註 20：Richard C. Bodner and Fei Song, "Knowledge-Based Approaches to Query Expansion in Information Retrieval" In Advances in Artificial Intelligence

(New York: Springer, 1996): 147.

- 註 21 : Hsinchun Chen, Tobun D. Ng, Joanne Martinez, and Bruce R. Schatz, "A Concept Space Approach to Addressing the Vocabulary Problem in Scientific Information Retrieval: An Experiment on the Worm Community System," Journal of the American Society for Information Science 48, no.1 (1997): 17-31.
- 註 22 : E. A. Fox, "Development of the CODER System: A Testbed for Artificial Intelligence Methods in Information Retrieval," Information Processing and Management 23, no.4(1987): 341-366.
- 註23 : M.E. Lesk, "Word-word Associations in Document Retrieval System," American Documentation 20, no.1(1969): 27-38.
- 註24 : H. Chen and D. T. Ng, "An Algorithmic Approach to Concept Exploration in a Large Knowledge Network (Automatic Thesaurus Consultation): Symbolic Branch-and-Bound vs. Connectionist Hopfield Net Activation," Journal of the American Society for Information Science 46, no.5(1995): 348-369.
- 註25 : H. Chen and V. Dhar, "Cognitive Process as a Basis for Intelligent Retrieval Systems Design," Information Processing and Management 27, no.5(1991): 405-432.
- 註26 : C. J. Crouch and B. Yang, "Experiments in Automatic Statistical Thesaurus Construction," In Proceedings of the 15th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, SIGIR 92 Copenhagen, Denmark, June 21-24 1992, 77-88.
- 註27 : A. T. McCray and W. T. Hole, "The scope and structure of the first version of the UMLS semantic network," In Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care, , Los Alamitos, CA: Institute of Electrical and Electronics Engineers, November, 4-7 1990, 126-130.
- 註28 : A. Y. Chamis, Vocabulary Control and Search Strategies in Online Searching, (New York: Greenwood Press, 1991), 128.

- 註29： S. D. Knapp, "Creating BRS/TERM, a Vocabulary Database for Searchers," DATABASE 7, no.4(1984):70-75.
- 註30： A. B. Piternick, "Searching Vocabularies: a Developing Category of Online Search Tools," Online Review 8, no.5(1984): 441-449.
- 註31： A. S. Pollitt, "A Front-End System: An Expert System as an Online Search Intermediary," ASLIB Proceedings 36(1984): 229-234.
- A. S. Pollitt, "CANSERCH: An Expert Systems Approach to Document Retrieval," Information Processing & Management 23(1987): 119-136.
- 註32： G. Brajnik, G. Guida and C. Tasso, "IR-NLI II: Applying Man-Machine Interaction and Artificial Intelligence Concepts to Information Retrieval," In Proceedings of the 11th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, SIGIR 88 Grenoble, France, 1988, 387-399.
- 註33： W. Bruce Croft and R. H. Thompson, "T³R: A New Approach to the Design of Document Retrieval Systems," Journal of American Society for Information Science 38(1987): 389-404.
- 註34： P. Shoval, "Principles, Procedures and Rules in an Expert Systems," Communications of the ACM 29(1985): 475-487.
- 註35： 陳昭珍，「主題索引問題初探」，ASIS台北學生分會會訊 第五期(民國81年6月)：頁29-31。
- 美國資訊科學學會台北分會編，索引典理論與實務(台北市：編者，民國83年)，頁86-93。
- 註36： 同註12，頁131。
- 註37： 黃卓倫，「利用隱藏語意索引進行文件分段檢索之研究」，國立台灣大學資訊管理研究所碩士論文，民國85年，頁46-47。
- 註38： 同註8。

註39： 同註37，頁47。

註 40： 同註 20，151-152.

伍、XML/Metalogy 管理系統之權威控制系統建置

一、數位圖書館系統架構

基本的數位化系統，從使用者的角度而言，應包含的功能主要有查詢、瀏覽及其他相關網路資源連結。而通常資料量大的數位圖書館系統都會以資料庫來管理其書目記錄及數位化物件、或連結的網址，為避免連結網址的失效，及數位化物件更改存放的位置，所以也會考慮使用 Handle system 之類的機制來做識別處理(Identifier)。此數位圖書館之基本架構如圖一所示：

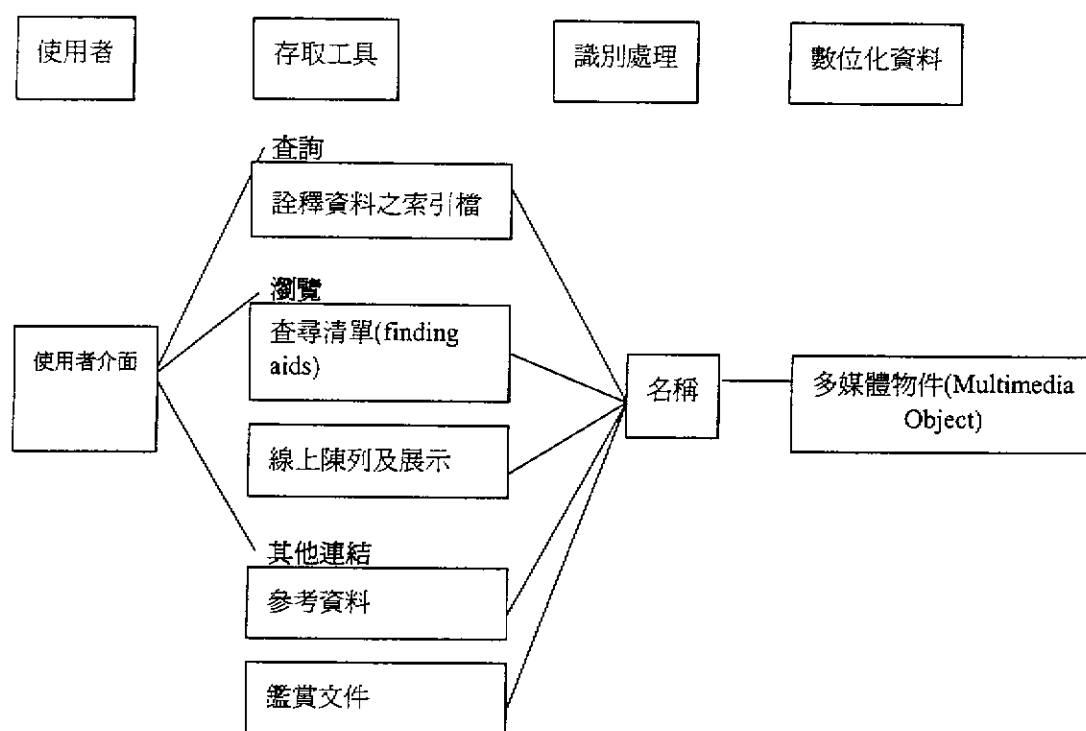


圖 12：數位圖書館系統架構

以目前的檢索技術而言，提供查詢的方法不外為：全文檢索或欄位查詢。全文檢索可以不必對資料做描述，但檢索效能較差，對於無文字的影像、聲音或視訊資料而言，在內容檢索技術(content-based retrieval)未成熟之前，全文檢索技術也無法派上用場，所以以人工描述 metadata，建立欄位化的書目資料，對於數位圖書館而言，是絕對必要的工作。因此，在一數位圖書館或博物館等系統中，Metadata 扮演重要的角色。Metadata 的訂定需由瞭解使用者需求及資訊屬性入

手，同時也需考慮到資訊系統之間的互通性，所以相關標準的採用非常重要，如國際間已發展出來的 Metadata 格式：Dublin Core、EAD、FGDC、GILS、TEI；以及 Metadata 的交換語法，如 SGML/XML/HTML 等。目前發展中的數位圖書館系統不管用那一種 metadata，其交換語法大多選用 XML 或 SGML，尤其是 XML 語法是可取 SGML 之長、補 HTML 之短，也是目前網際網路界極力推廣的語法。本文主要在說明設計 Metadata 時所涉及之問題，並介紹我們所發展的一套 XML/Metadata 通用系統 – Metalogy 之特色、架構、功能及使用方式。

二、Metadata 之意義與功能

(一) Metadata 之意義

所謂 metadata，在資訊組織界最普遍的解釋是 “data about data”，意指有關資料的資料，即資料之描述性資訊，如圖書館的 MARC 記錄，即為一種 metadata。但 Renato Innella 認為這種解釋，忽略了 metadata 最重要的概念，即其結構性，他認為應將 metadata 定義為 “structure data about data” 此結構二字，使得採用 metadata 做組織資訊的方式和全文索引 (full-text indexing) 有所區隔。(註 1) 對資料做描述以建立索引的作法並非始於今日，但 metadata 一詞卻是在網路資訊出現後，才普遍被使用的，且多指網路資源等電子資訊的描述。國內翻譯名詞有陳雪華等的詮釋資料 (註 2)，吳政叡之“元資料” (註 3)，筆者也曾將之譯為“超資料” (註 4)，紛歧顯見。

(二) Metadata 之功能

Locan Dempsey 認為 metadata 的功能主要有：(註 5)

1. 找到資訊之位址 (location)；
2. 蒐尋資訊 (discovery)；
3. 紀錄資訊 (documentation)；

4. 評估資訊 (evaluation);
5. 選擇資訊(selection);
6. 其他。

Renato Iannella 及 Andrew Waugh 也提出類似的觀點，他們認為 metadata 主要用來：(註 6)

1. 概述資料之內涵
2. 讓使用者查到該資料
3. 讓使用者決定該資料是否是他所要的
4. 避免使用者存取該資料（如禁止兒童使用某類資料）
5. 讓使用者檢索、複製資料
6. 指示應如何解譯該資料（例如說明資料之格式、編碼、加密的情況）
7. 用來決定可檢索哪一個資料（若資料可多種格式存在）
8. 說明資料使用的合法情況
9. 說明資料之歷史，如說明其原始資料為何，及其他的改變
10. 說明資料之聯絡人，如擁有者
11. 指示該資料與其他資源的關係
12. 控制資料的管理

但在資料庫管理界，metadata 則直指資料庫管理系統之架構(scheme)，且其對 metadata 的討論偏重在如何藉由 metadata 的使用，以達到多資料庫系統(multi-databases)的互通(interoperability)等問題。香港科技大學的 Pamela Drew 及 Jerry Ying 認為傳統的 metadata 被定義為：有關資訊源的靜態觀點(a

static view over some source information), 此觀點通常由 metadata 應如何讓使用者蒐尋及存取資訊源之角度來思考。但是現在, 因整個廣域網路中資訊源需互連, 故而傳統 metadata 的範圍與角色也隨之改變。所以, 他們認為, 現在的 metadata 應滿足三項需求: (註 7)

1. 促使系統互通, 而不只是在提供摘要性資訊;
2. 當越來越多的資訊被電子化時, metadata 模組應能讓電腦連接資訊源並自動擷取 metadata;
3. metadata 管理系統應能定期核對原始資訊源, 以確保 metadata 資訊的正確性。

羅森寶(Rothenberg, Jeff)認為和一資料庫有關之 metadata 有三類: (註 8)

1. 協助或限制檢索的 metadata;
2. 作為分享與互通之用的 metadata;
3. 表達資料特性及作為索引資料的 metadata。

同時羅森寶更進一步的將 metadata 分為三種層次, 每一層中所包含之資訊如下:

1. 資料庫層次之 metadata(database level metadata)
2. 資料項目層之 metadata(data-element level metadata)
3. 資料值層之 metadata(data-value level metadata)

(三) Metadata 相關之課題

如果將羅森寶的三種層次再進一步詮釋，我們也可以說 Metadata 的設計必需考慮三方面的問題，一是語意(semantic)、二是結構(structure)、三是語法(syntax)。這三者的關係如下圖所示：

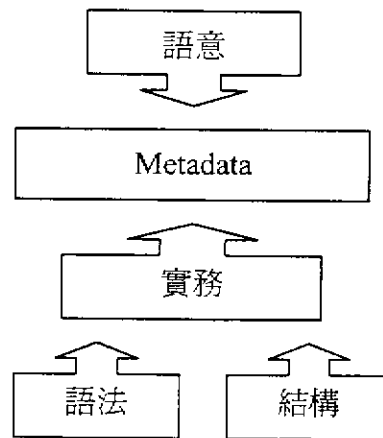


圖 13：Metadata 的三個問題

此外，分散式資訊的整合檢索雖不是 Metadata 的直接內涵，但在網路環境中是一個重要的課題，而且相關標準的設計和 Metadata 習習相關。茲分別說明如下：

1. 語意問題 (semantics):

Metadata的語意內涵乃根據資料的特性及使用者的需求而定，它探討的問題包括應該描述資料的那些屬性？屬性的名稱如何命名？描述該屬性時所根據的資訊來源為何？以及該屬性內容是否應根據索引點或權威檔來描述？AACR2、Dublin Core、AAT (Art and Architecture Thesaurus)、LCSH (Library of Congress Subject Heading)、LCNA (Library of Congress Name Authority File) 等屬於此類標準。

2. 結構問題 (structure):

決定了資料應描述屬性後，接下來的問題是建立屬性的結構。

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、不疊床架屋，且便於程式的撰寫。MARC、DTD、RDF 等所關心的即是這個層次的問題。

3. 語法問題 (syntax):

要使 Metadata 具備互通性，可以在不同系統之間交換，則包裝 metadata 的語法是非常重要的部份。猶如 ISO 2709 在不同的圖書館自動化系統之間穿梭自如，HTML 是使得 WWW 文獻能在不同系統間交換的主要功臣、而 SGML、XML 是電子圖書館/博物館/檔案館系統用來標示其 Metadata 及全文資料的標準語法。

4. 整合檢索的問題

即使每個系統都根據相同的 Metadata 描述資料，但各系統的索引模式與檢索指令仍會不同，因此使用者往往必需學習不同系統的檢索介面及指令以查尋不同的系統。當資料庫很少的時後，這樣的學習還可忍受，若資料庫很多時，太多檢索指令及介面的學習則毫無效率可言。更何況在電子圖書館/博物館/檔案館領域中，不同學科領域或資訊類型往往使用不同的 Metadata。因此分散式資訊的整合檢索在網路環境中是一個非常重要的課題。Z39.50、START 等即屬此類的標準。

三、Metadata 的語法—SGML 與 XML

無論就數位圖書館相關計畫或電子商務等網際網路深入的應用而言，XML 已成 Internet 界積極推廣的語法。XML 是 W3C 在 1996 年底提出的標準，它是從 SGML 衍生出來的簡化格式，也是一種元語言 (meta-language)，可以用來定義任何一種新的標示語言。XML 的制定是為了補足 HTML 的不完美，使得在 Web 上能夠傳輸、處理各類複雜的文件，它去除了 SGML 複雜不常用及不利於在 Web 傳送的選項功能，讓使用者可以很容易地定義屬於自己的文件型態，程式設計師也能在更短的時間開發 XML 相關應用程式 (註 9)。XML1.0 Recommendation 已於 1998 年 2 月公佈，相關標準目前仍在發展之中，XML 的

發展獲得了各界的支持，其中包括了 Sun Microsystems, Microsoft, Netscape, Adobe, ArborText...等軟體大廠的背書（註 10）。XML 與 HTML 及 SGML 到底有何不同呢？

（一）XML 與 HTML 之比較：

XML 與 HTML 是不同層級的標示技術。HTML 是一種特定的標示語言，包含了一組標準的元素集（element set），元素被用來指示特定的文字片段在文件中所扮演的角色，每一個元素都有標準的意義。例如：<H1>與</H1>在 HTML 文件中的意義是最重要的標題（heading）（註 11）。其實除了 HTML 之外，還有很多種特定的標示語言，例如航空產業針對飛行器維護手冊制定了一個標示語言稱為 ATA-2100，半導體產業針對電路資料制定了一個標示語言稱為 PCIS，電腦產業針對軟體文件訂了一個標示語言稱為 DocBook。這些個別的標示語言都有以下共同點：（註 12）

- 都定義了一組標籤集（tag set），每個標籤都有標準的意義及標準的使用規則，也就是有標準的文法（grammar）。
- 都是為了某類型的文件或資料專門設計的。
- 都使用文字處理（text processing）國際標準 SGML 來定義它們的標籤集與文法。

所以 HTML 與 PCIS 及 DocBook 一樣，都是 SGML 的一種應用，只不過 HTML 所強調其超連結及網頁呈現之功能，且為這些功能寫出一組固定的標籤集。這也就是 XML 與 HTML 主要的差異，茲再詳細說明如下

1. XML 與 HTML 在特性上主要的不同：

XML 繼承了 SGML 的可擴展性（extensibility）、結構性（structure）與可確認性（validation），這三點也是 XML 與 HTML 在特性上最重要的不同（註 13），詳細說明請見表一。

表 1：XML 與 HTML 在特性上主要的不同

	HTML	XML
可擴展性 (extensibility)	標籤集與每個標籤的意義是固定的，使用者無法自行定義標籤（tags）或屬性（attribute）。	允許使用者根據需要，自行定義標籤與屬性，以便更進一步描述資料的語意。
結構性 (structure)	不支援深層的結構描述。	能描述各種複雜的文件結構，並能表示資料庫綱要（schema）及物件導向的階層結構。
可確認性 (validation)	沒有提供語法規格來支援應用程式對 HTML 文件進行結構確認。	可包含一個語法規格（DTD），讓應用程式對文件進行結構確認。

2. XML 與 HTML 在語法上主要的不同（註 14）

XML 與 HTML 在語法上的差異主要有：

(1) XML 文件必須遵守階層的元素結構(hierarchical element structure)：

XML 文件每一個起始標籤都要有對應的結束標籤，成對的起始標籤與結束標籤就稱為元素。每一元素都必須巢狀包覆（nested within）在另一個元素之內。例如：

```
<LI>HTML allows <B><I>improper nesting</B></I>
```

上面是一段 HTML 原始碼，但 B、I 並沒有巢狀包覆的關係，且 LI 沒有結束標籤，所以不符合 XML 的語法規則。

```
<LI>XML required <B><I>proper nesting</I></B>
```

上面這個例子則符合 XML 的語法規則，I 巢狀包覆在 B 之內，且每一個元素都有成對的起始標籤與結束標籤。

(2) XML 與 HTML 對於空元素（Empty elements）的表示方式不同：

和 HTML 一樣，XML 也允許元素是空元素，所謂空元素的意思是元素的起始標籤與結束標籤是同一個。XML 對於空元素的表示方法和 HTML

不同，必須在元素型別名稱之後加上斜線(slash)。例如：下面這一段 HTML 原始碼中，
是一個空元素。

```
<LI>This is a HTML <BR>empty tag.</LI>
```

以 XML 的語法規則，必須表示如下：

```
<LI> This is a XML <BR/>empty tag.</LI>
```

XML 這樣規定的目的是使 XML 處理器 (processor) 讀到 “/” 便知道這是一個空元素，就不用再往下找結束標籤了。在 HTML 中，HTML 處理器之所以知道
 是一個空元素，是因為 HTML 的標籤集是預先定義好的；而 XML 允許使用者自行定義標籤，因此必須有特殊的語法來表示空元素，XML 處理器才不至於搞混。

(3) XML 所有的屬性值 (attribute value) 必須以引號 (quotes) 夾住：

HTML 除了包含空白的屬性值必須以引號夾住外，其餘的情況可以省略引號；而 XML 所有的屬性值 (attribute value) 都必須以引號 (quotes) 夾住。例如：

```
<FONT color=#FF0000>HTML attribute value</FONT>
```

以上是一段合法的 HTML 原始碼，屬性值沒有用引號夾住，不符合 XML 的語法規則。

```
<FONT color="FF0000">XML attribute value</FONT>
```

上面這個例子符合 XML 的語法規則，屬性值必須用單引號或雙引號夾住。

(4) XML 標籤大小寫有區別 (case-sensitive)：

HTML 標籤大小寫是一樣的，XML 標籤大小寫則有區別。例如：

```
<H1>HTML is not case-sensitive, XML is.</h1>
```

以上是一段合法的 HTML 原始碼，在 HTML 中，H1 與 h1 是相同的標籤；但在 XML 中，H1 與 h1 則是不同的標籤。上面的 HTML 原始碼若

以 XML 的語法來寫，就得寫成

```
<H1>HTML is not case-sensitive, XML is.</H1>
```

(5) XML 與 HTML 對空白 (white space) 的處理方式不同：

HTML除了PRE元素內的連續空白有效外，其餘的連續空白都會被視為只有一個空白字元。HTML的空白包括了ASCII space()、ASCII tab(	)、ASCII form feed()、Zero-width space(​)以及 Line breaks，這裡的 &#x 表示Unicode的十六進位。(註15)而由於XML主要是用來裝載資料，因此在XML標籤之間的空白都是有效的，但在標籤本身及屬性值的引號之間的空白是無效的。例如：

```
(a)<desc lan="English">XML white space handling</desc>
```

```
(b)<desc
```

```
lan="English">XML white space handling</desc>
```

```
(c)<desc lan="English">XML   white space   handling</desc>
```

以上三段 XML 原始碼，(a)與(b)是相同的，但(c)與(a)(b)則不同。

(6) XML 文件只能有一個根元素 (root element)：

這個限制使得 XML 剖析器較容易確認文件已經結束了。

(7) XML 所有的實體 (entities) 必須在 DTD 中宣告：

實體的作用就好像程式語言中的常數 (constant) 一般，它會在文件處理時展開，就好像處理巨集 (macro) 的情況一樣。當文件中有許多重複的文字片段，便可以宣告實體來代替這些重複的文字片段，除了可避免人工重複輸入時發生錯誤，也使文件更簡潔與日後更容易維護。

(8) XML 允許指定不同的字集編碼 (character set encoding)：

XML 的字集編碼是在<?xml?>宣告中指定，例如：

```
<?xml version="1.0" encoding="UTF-8">
```

上面的例子表示這份 XML 文件以 UTF-8 編碼。所有符合 XML 規格的軟體工具一定都支援 ISO 10646 字集，ISO 10646 是一組龐大的字集，其中包含了 Big5 中文碼及 GB2312 中文碼中所有的字。ISO 10646 包含了幾種編碼方式：UTF-8 是 8 bit 的，UTF-16 是 16-bit 的，Unicode 就是 UTF-16 的形式。(註 16)

(9) XML 的特別保留字元 (special reserved characters)：

有一些字元本身就是 XML 語法結構的一部份，若要在 XML 文件中使用這些字元必須使用替代的表示法，XML 使用預先定義的實體 (predefined entity) 來替代這些字元，如表二所示：

表 2：XML 的特別保留字元

Reserved character	Predefined entity to use instead
<	<
>	>
&	&
“	"
‘	'

上表中前四個字元在 HTML 中也使用同樣的替代表示法，在 HTML 中稱為字元實體參引 (character entity references) (註 17)。

(二) XML 與 SGML 之比較

XML 和 SGML 都是元語言 (meta-language)，可以用來定義各種不同的標示語言。XML 可說是 SGML 的簡化版本，繼承了 SGML 的可擴展性 (extensibility)、結構性 (structure) 與可確認性 (validation)，但去掉了 SGML 中造成應用程式不易開發的特徵。XML 與 SGML 的相同點及相異點如表三所示：

表 3：XML 與 SGML 的相同點

相同點	● 都是一種 Meta-Language，可以支援各種不同的應用。
	● 都具備可擴展性(extensibility)、結構性(structure)與可確認性(validation)等優點，能夠描述各種複雜的文件結構，並可藉由 DTD 來確保文件結構的完整性與正確性。
	● 都必須結合樣式表來設定文件內容的呈現格式。
	● 都具備跨平台、可長期保存、可再利用、能被人直接閱讀等特性。
主要相異點	➢ 對 SGML 文件而言，DTD 是必備的；而對於 XML 文件而言，DTD 不是必備的。Well-formed XML 文件沒有 DTD，Valid XML 文件則有 DTD。
	➢ 所有的 XML 文件都必須符合下列 Well-formedness 規則： ✓ 包含一個以上的元素 (elements)。 ✓ 恰有一個根元素 (root)，不出現在其他元素的內容之中。 ✓ 所有的元素都有起始標籤與結束標籤，也就是標籤必須成對。 ✓ 所有的標籤必須呈現適當的巢狀 (nest) 結構。 ✓ 空標籤 (empty tags) 必須遵守特殊的 XML 語法。(如<empty/>) ✓ 所有的屬性值都必須括上單引號或雙引號。 ✓ 所有的實體都必須宣告。
	➢ SGML 的使用者可以使用包含 (Inclusion) 與除外 (Exclusion) 這兩種語法規則來指定內容模式 (content model) 的例外處理 (Exceptions)；XML 則不支援例外處理。
	➢ XML 不支援 AND(&)內容模式、SDATA 內部實體 (internal entities) ... 等語法。

XML 與 SGML 之間除了上表主要差異點外，在語法上還有很多細部的差異。SGML 在語法上具有相當大的彈性，使用者可以針對不同的需要選擇不同的特徵，但語法上的彈性愈大，相對也使得開發 SGML 剖析器（parser）與應用程式的難度及所需投入的成本提高很多，這是 SGML 無法被普遍接受的主因之一。鑒於 SGML 在發展上遭遇的限制，XML 在設計上採取不同的策略，XML 捨去了 SGML 中相當多的選項功能，藉著在語法上加入較多的規定與限制，來換取開發 XML 剖析器與應用程式的簡單性。要真正了解 XML 與 SGML 細部的不同，必須對 SGML 語法相當熟稔，對於原本早已使用 SGML 來處理電子文件的單位，若要轉換到 XML 當然必須研究它們之間在語法上差異，進而訂出轉換規則；而對於剛要採用 XML 的單位事實上只需以 XML 為起點即可。

四、XML/Metadata 系統發展實例

由上述說明可知，XML 具有 SGML 的特性，但又不似 SGML 那麼複雜；XML 便於在網路上傳輸，又可以加強 HTML 所沒有的彈性及精確性。所以它已成為網路界及資料庫界積極支持的語法。因此，當我們要設計一個 Metadata 管理系統時，也決定採用這種語法，做為資料庫間交換資料的主要依據。但是除了語法外，如前所述，尚有語意的問題，由於目前已發展出來的 Metadata 格式相當的多，此外，為了能表達各類型各主題資料之屬性，尚未發展 metadata 格式的領域可能還會發展出該領域的 metadata 格式，所以一個有彈性的 metadata 管理系統，不能只是針對一種 metadata 格式來開發，而應該能讓使用單位自己決定要用那一種格式。所以，開發一個 XML/Metadata 的通用性系統正好可以滿足這種需求，這也是 Metalogy 這個系統最主要的特色。詳細設計理念及架構說明如下：

（一）Metalogy 系統特色與架構

Metalogy 是我們在國科會數位博物館專案下（註 18），費時一年多所發展出來的 XML/Metadata 管理系統，這個系統可以用來建立數位博物館、數位圖書館、數位檔案館或任何主題資料之資料庫。它具有利用 DTD 建立資料庫、Metadata

編輯、權威檔(或索引典)編輯、查詢(含 Window 介面及 Web 介面)、及轉入及轉出 XML 記錄等功能，此系統之特色主要有下列數端：

- a. 系統綱要(schema)主要依據輸入之 DTD 決定。
- b. 系統允許兼容多種不同 DTD 同時存在。
- c. 多種不同格式之資料能夠同時被檢索。
- d. 資料處理之結構允許使用者依據綱要調整內定之欄位格式與使用條件。
- e. 具有親和的介面讓使用者自訂超連結、索引、檢索及顯示欄位。
- f. 資料轉入與轉出能依循所屬之 DTD 格式。
- g. 系統能判定轉入之資料是否符合指定之 DTD 格式，以及執行重複查核。
- h. 資料內部處理能包含欄位式結構、多媒體結構、與全文。
- i. 系統包含存取控制(access control)、異動記錄(transaction log)等系統管理功能。
- j. 另有 Web 查尋軟體，可讓一般使用者透過 WWW 查詢資料庫。

此系統之架構主要如下圖所示：

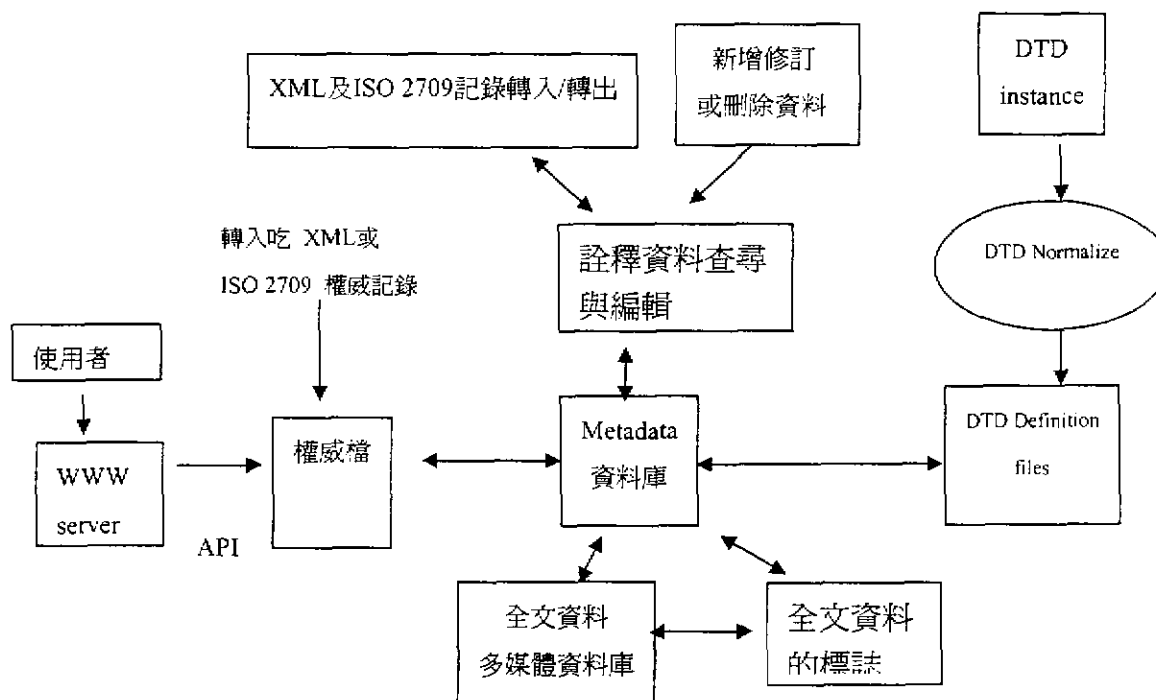


圖 14：Metalogy 系統架構圖

(二) 系統檔案

Metalogy 系統檔案包括下列各項：

1. 「STAFF」使用者檔—記錄各工作人員之基本資料，與其操作之權限。主要用途為登錄系統。登入時必須輸入使用者代碼與密碼，完成確認時系即會記錄此一使用者允許執行之權限。當選則系統任一功能時，便一據此一權限比對該功能之操作權限半段是否允許操作。若有兩人以上同時擁有允許操作之權限時，卻需限定只有一位能執行該功能的情形，可考慮將該功能設定「指定密碼」，也就是說，當執行該功能時，使用者之權限必須大於該功能之操作權限，且必須輸入正確之密碼方可執行。統預設為使用者為「TEST」，密碼亦為「TEST」。
2. 「BRANCH」使用單位/館別檔—記錄各使用單位之基本資料。系統允許多館 (multi-branch) 作業。
3. 「OPERATION_LOG」系統異動記錄檔—主要記錄系統異常之情況，提供爾後客服人員檢視、判斷異常發生原因與時間。另外亦可記錄特定處理功能之

執行狀況，提供系統管理人原監督某些特定功能之使用狀況與追蹤何人曾經操作。

4. 「FUNCTION」功能操作權限檔—設定各操作功能之允過權限、密碼、與操作時是否記錄至 OPERATION_LOG 檔案。
5. 「MSG_ERR」錯誤訊息檔—指定各錯誤訊息之說明內容、圖示、與按鈕種類。
主要目的是提供系統在不須重新編譯的情況下，能夠隨時一使用者實際需要修改顯示的訊息內容，以避免部份訊息用語含糊或術語不夠專業，造成使用者誤解。
6. 「SYS_LOCK」系統鎖檔—多人執行編輯資料時，管制使用者之系統鎖定記錄。當使用者維護某比資料時，系統會自動將該資料編號與操作人員代碼、時間記錄於此檔案內，以防止其他人員同時維護此筆資料，維護完畢細懂會自動清除資料之鎖定記錄。但當維護資料途中，發生系統當機或操作人員之電腦當機、網路突然中斷等情形時，系統無法自動清除該資料的鎖定記錄，在此情形下必須經由人工解除鎖定。
7. 「LIST_CODE」代碼檔名細檔—記錄系統內部所有使用者自訂之代碼檔名細(lists)。系統內部所有以「CODE_」開頭之檔案，均代表「使用者自訂代碼檔」，例如 CODE_LANG 代表語文代碼；CODE_EDU 代表學歷代碼。系統允許使用者自行新增代碼檔，與自行修改代碼黨內之代碼資料。

【注意】代碼檔案內所設定之代碼與其含意(說明)只允許新增與修改，不允許刪除，主要是為了避免資料對應問題。因此在新增代碼時必許注意此一問題。

8. 「PARAMS」系統流水號檔—系統各種資料存檔所需之流水編號均記錄於此檔案內。不同資料所需之流水號碼預設起始為 1，但為因應部份資料必須可慮編號格式可事先設定起始值(例如登錄號為 6 為數字，則設定起始值為 100001)，因為流水號為數字格式，因此無法提供「000001」這一類前置為 0 的格式。

(三) 編目參數檔案

資料檔案關連圖(尚未包含權威控制檔)

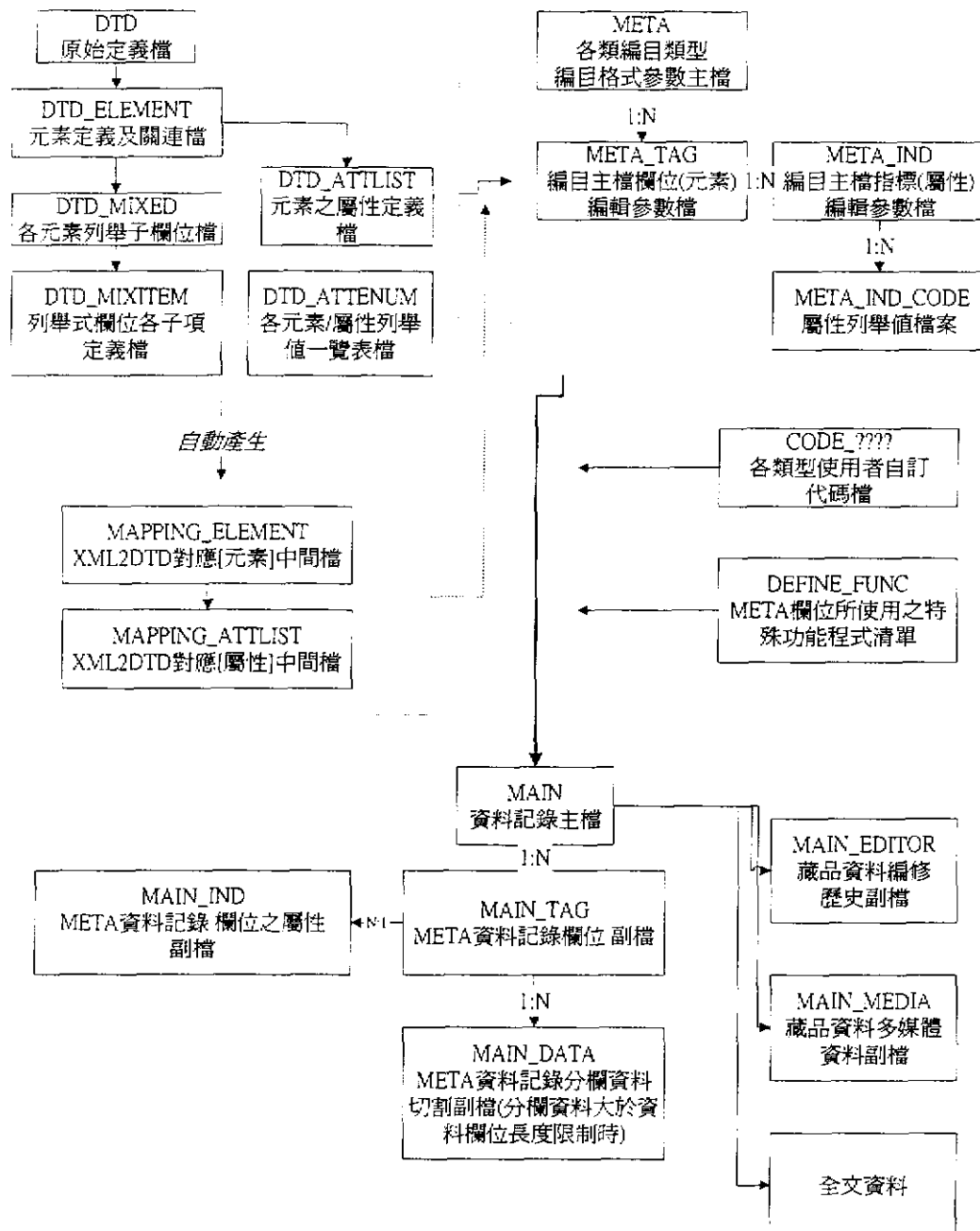


圖 15：資料庫檔案關連圖（未包含權威控制檔）

1. 因為 Metalogy 系統主要資料結構的綱要是來自 DTD，也就是 Metadata 的依據。因此資料庫內部必須能夠儲存轉入之 DTD 資訊。由於 DTD 是屬於巢狀結構，類似於物件導向資料庫表格。而 Metalogy 為簡化系統前端的軟體開發複雜度與後端資料庫透通性，採用關連式資料表格做為主結構。因此一個 DTD 檔案轉入至 Metalogy 系統時，依據 DTD 元素的特性必須分別儲存於下列各表格內：
 - (a)DTD：記錄原始 DTD 轉入時的資訊，例如時間、說明、以及原始內容等。
 - (b)DTD_ELEMENT：記錄各 DTD 所有之元素名稱、與其上下層關係。
 - (c)DTD_ATTLIST：記錄各元素所屬的屬性集
 - (d)DTD_ATTENUM：記錄各屬性的列舉值內容。也就是各屬性允許輸入的代碼表列。
2. 「META」一因為 DTD 在定義上並不限定根元素之名稱，且每筆 XML 均各自獨立一個檔案。因此在轉入 DTD 之前必須事先告知系統，該 DTD 的名稱、系統處理時的控制代碼、以及轉出 XML 資料時的封裝根元素名稱。如此便可達到：不同 DTD 能避免相同根元素/元素名稱的困擾；能將批次轉出的多筆 XML 資料封裝在一個指定的「封裝根元素」之內，反之轉入時也可依循相同模式處理。因此，若要檢視系統內具備的 metadata 種類，最直接的方式便是查閱 META 表格。
3. 由於 DTD 在宣告上並不具備資料格式、特殊處理轉換、輸入長度、所屬權威款目、索引種類等欄位定義。另外使用者在處理資料時也可能希望將轉入之 DTD 做適度轉換，例如合併元素，或只採用部份元素等等。基於上述考量，系統必須另外存在一組「映對」(mapping)表格，作為將 DTD 檔案定義轉化成 Metalogy 系統內部 meta 結構的中間檔，也就是形成 Metalogy 綱要的主要依據。

(a)MAPPING_ELEMENT：記錄各 DTD 所有之元素名稱、與其上下層關係。

(b)MAPPING_ATTLIST：記錄各元素所屬的屬性集

【附註】映對表格可以在轉入 DTD 時由系統依照該 DTD 內容自動產生，但仍需使用者人工檢視各元素的資料格式、特殊處理轉換、輸入長度、所屬權威款目、索引種類等欄位定義等等。另外也可事先編輯好，在執行轉入 DTD 的作業。

4.透過系統將 DTD 原始定義與映對表格比對結果，即產生 Metalogy 系統內部的 meta 結構，也就是作為該類型資料編輯的依據。這些資料分別存存於：

(a)META_TAG：【meta 欄位定義】記錄各 meta 所有之元素名稱、與其上下層關係。

(b)META_IND：【meta 指標定義】記錄各元素所屬的屬性集

(c)META_IND_CODE：【meta 指標列舉值】記錄元 DTD 屬性的列舉值內容，作為屬性指定輸入的代碼表列。

【附註】除原始 DTD 定義所指定之列舉值之外，許多欄位有其輸入之代碼表列。因此若 DTD 對於該欄位並無定義列舉值時，並需透過「映對表格」指定其所屬的代碼檔，Metalogy 系統內部允許使用的代碼檔請參閱「系統檔案」LIST_CODE 說明。

5. 「DEFINE_FUNC」—特殊檢查程式，例如日期格式、年代轉換、有效格式判斷等特殊程序之設定。由於牽涉特殊的程式邏輯，因此使用者不能自行設定，必須透過系統版本更新(patch)方式增加。

(四) 編目資料檔案

編目參數檔、資料檔與權威檔之關係

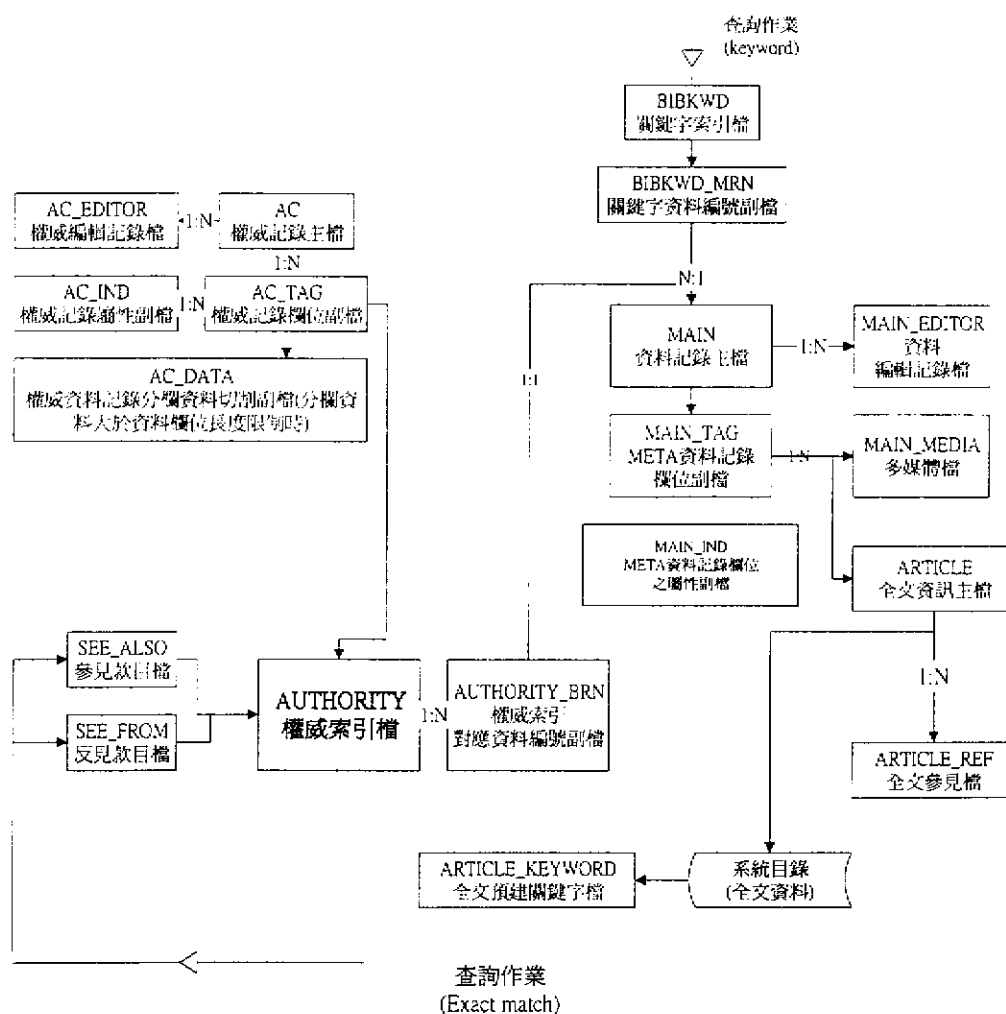


圖 16：編目參數檔、資料檔與權威檔之關係

編目資料流程圖

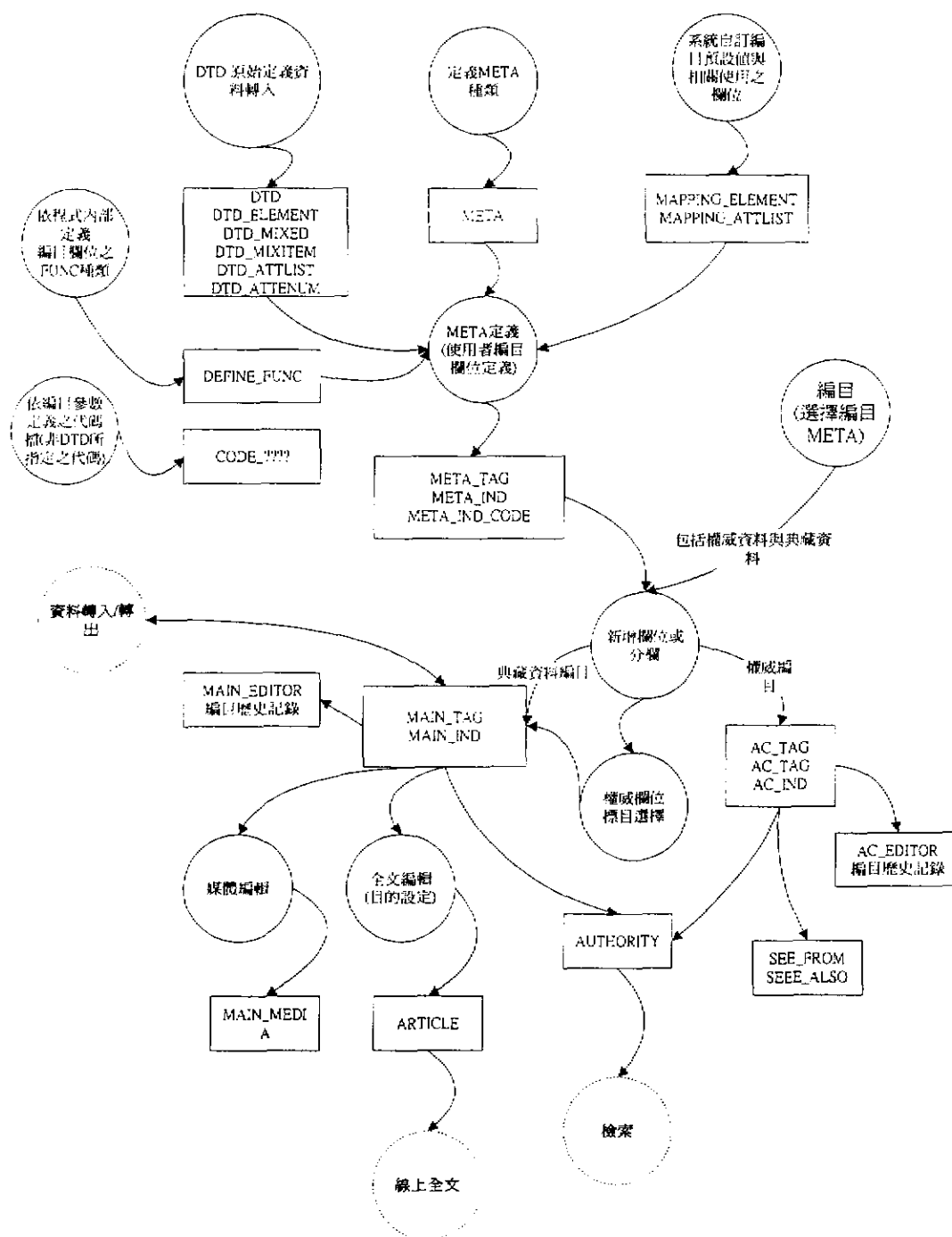


圖 17：編目資料流程圖

1. 所有典藏資料均儲存於 MAIN 表格之內，藉由其中 META 欄位區辨該筆記錄所屬的 meta 種類，也就是原 DTD 名稱：

(a)MAIN：典藏資料主檔。系統編號依據 PARAMS 表格之「LASTMRN」流水號作為各筆典藏紀錄之系統編號。當新編完成一筆記錄並存檔時，或新轉入一筆資料成功時，系統即會將該筆記錄予以編上此系統號。

(b)MAIN_EDITOR：紀錄每筆記錄的異動時間與操作人員等資訊。

(c)MAIN_TAG：儲存欄位的資料。相對應於 XML 的元素(element)。

(d)MAIN_IND：儲存指標的資料。相對應於 XML 的屬性(attribute)。

(e)MAIN_MEDIA：儲存多媒體資料，與 MAIN_TAG 表格之關係為多對多 (M:N)關係。[參見操作說明]

2. 所有權威資料均儲存 AC 表格之內，藉由其中 META 欄位區辨該筆權威記錄所屬的 meta 種類，也就是原 DTD 名稱：

(a)AC：權威資料主檔。系統編號依據 PARAMS 表格之「LASTARN」流水號作為各筆權威紀錄之系統編號。

(b)AC_EDITOR：紀錄每筆權威記錄的異動時間與操作人員等資訊。

(c)AC_TAG：儲存欄位的資料。相對應於 XML 的元素(element)。

(d)AC_IND：儲存指標的資料。相對應於 XML 的屬性(attribute)。

3.「AUTHORITY」權威索引檔—當新增一筆典藏資料的欄位，若其欄位為主款目(或任何非反見款目之檢索欄位)時，系統便會自動產生一筆權威索引檔。

(a)若該筆資料為反見款目，則會另外在「SEE_FROM」表格產生一筆記錄，標示此為反見款目。

(b)若該筆資料為參見款目，則會另外在「SEE_ALSO」表格產生一筆記錄，標示此為參見款目。

【附註】系統允許在尚未建立權威紀錄之前，即進行編目。編輯之欄位屬於主款目？參見款目？亦或是反見款目？系統是依據 MATA_TAG 編目參數檔的參數所設定。

當事後再建立權威紀錄時，系統亦可依據「AUTHORITY」、「SEE_FROM」、「SEE_ALSO」紀錄內容將權威紀錄的連結關係建立起來。

(五) 查詢參數檔案

查詢功能資料流程圖

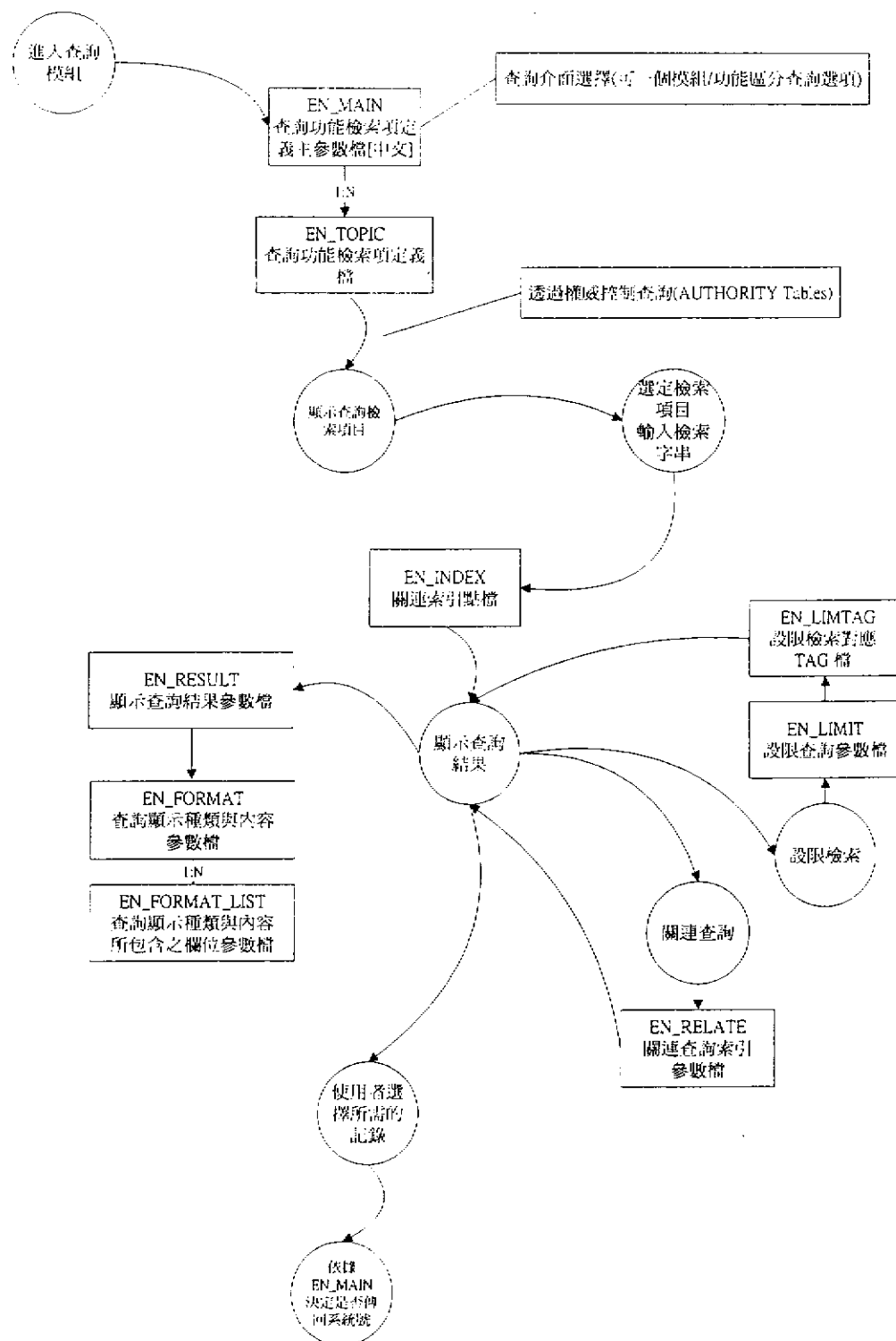


圖 18：查詢功能資料流程圖

1. 「EN_MAIN」查詢功能檢索項定義主參數檔—查詢功能之最基本檔案。決定查詢功能之基本需求，包括檢索種類(對應相關聯之 EN_TOPIC 檔)、使用者或是工作人員操作之規範、可否下載查詢結果之資料等等。因此若需在不同操作模組提供不同之檢索功能與檢索項目等，便必須設定個別之 EN_MAIN 資料。
2. 「EN_TOPIC」查詢功能檢索項定義檔—檢索項目(access pointer)之定義參數檔。定義個別檢索項目之預設查詢方式(關鍵字、切截、完全符合查詢等)，並可藉此檔案設定查詢之上下限筆數。

—查詢結果超過上限筆數，只顯示上限筆數內之資料數量。

上限

—查詢結果之筆數介於上下限筆數之間，另提供設限查詢。

下限

—查詢結果之筆數不超過下限之筆數時，隱藏設限查詢。

【附註】預設上限設定為 200，下限設定為 20。設限之欄位種類，必須依據「EN_LIMIT」檔案與其附檔「EN_LIMTAG」決定。

3. 「EN_INDEX」關連索引點檔—實際檢索典藏資料各欄位之索引點。例如 EN_TOPIC 檔案可設定一檢索項目為「地點」，但透過 EN_INDEX 檔案設定「地點」實際為檢索「事件地名」與「使用地點」等典藏資料的欄位。而實際檢索的欄位種類又可依不同之檢索需求(依據 EN_MAIN)而設定。
4. 「EN_FORMAT」查詢顯示種類與內容參數檔—顯示查詢結果之資料欄位內容。顯示之內容包括資料欄位之說明與欄位之內容。系統包括三組顯示格式

—「FULL」完整顯示格式、「MEDIUL」一般顯示格式、「BRIEF」簡略顯示格式。系統管理者可依據需要，制定此三種格式之顯示欄位內容。

5. 「EN_RESULT」顯示查詢結果參數檔—和「EN_FORMAT」顯示格式最大之不同是「EN_RESULT」是提供查詢第一層之顯示內容，僅具備單行之顯示內容。當使用者選擇某一檢索項目進行檢索時，所顯示之最初查詢結果及是依據「EN_RESULT」檔而決定，當使用者選擇其中某些資料欲查看詳細內容時，才是依據「EN_FORMAT」檔來決定呈現之內容。
6. 「EN_RELATE」關連查詢索引參數檔—配合「EN_FORMAT」顯示資料時，在指定之欄位上顯示超連結之訊息，提供使用者在檢視該筆資料內容時，可以關連再檢索出相關的資料。
7. 「EN_LIMIT」設限查詢參數檔—包括檢索結果第一層(EN_RESULT)與檢索第二層(EN_FORMAT)均可提供設限之功能。為考慮系統處理負荷與網路傳輸資源，必須限制查詢之筆數，因此在符合上下限檢索範圍內之筆數時，提供設限功能供使用者收斂查詢之筆數。一般而言，年代、種類、地點或屬於代碼項目之欄位均非常適合作為設限檢索之欄位。
8. 「BIBKWD」關鍵字主檔—一般關鍵字詞檔。負責控制關鍵字與同音、西文之單複數等容錯之檢索功能。

(六) 系統開發工具

Metalogy應用系統所用之程式語言為Delphi 5.0， Web Searching 所用之程式語言為ASP，後端可用的資料庫管理系統為Oracle 及 SQL server。

(七) Metalogy 系統主要功能

目前 Metalogy 已開發出來的功能包括：

1. 載入 DTD，建立資料庫
2. 定義資料庫綱要(schema)
3. metadata 編目功能
4. 數位化資料的管理與描述
5. 查詢功能
6. 載入 XML 記錄
7. 轉出 XML 記錄
8. 使用者權限控制
9. 訊息管理
10. Web 查詢功能

(八) Metalogy 操作手冊

為了便於使用者利用本系統，我們也製作了 Metalogy 操作手冊，以圖示法詳細說明本系統的安裝、設定、使用及操作方法，相信透過本手冊，使用單位可以不假他人的幫助，輕易的就能使用自如。關於 Metalogy 的操作說明請見子計畫三。

(九) Metadata 及 DTD 範例

設計數位博物館/圖書館/檔案館之際，一定會針對要處理的資料設計 Metadata，而要使用 Metalogy，也一定要將 Metadata 以 XML/DTD 來表示。由於 Metadata 及 DTD 的訂定需要對使用者之資訊需求及所處理的資訊屬性深入了解，相當費時，也需考慮系統的互通性，最好能使用標準。我們今年為故宮博物院的數位博物館計畫設計了幾份 Metadata 及 DTD，包括：書畫 metadata 及 DTD、器物 metadata 及 DTD、文獻 metadata 及 DTD、展覽 metadata 及 DTD、參考書目之 metadata 及 DTD、人名權威檔之 metadata 及 DTD、主題權威檔之 metadata 及 DTD、地名索引典之 metadata 及 DTD、時代索引典之 metadata 及 DTD，同時也有著錄手冊及著錄範例。請參考子計畫一。

註釋：

- 註 1：Renato Iannella, "Mostly Metadata: A Bit Smarter Technology," <http://www.dstc.edu.au/RDU/reports/VALA1998/>
- 註 2：陳雪華，陳昭珍，陳光華，「數位圖書館/博物館中詮釋資料之理論與實作」
圖書館學刊，第十三期（民國 87 年 12 月）：頁 37-59
- 註 3：吳政叡，"從電子檔案和元資料看未來資料著錄的發展趨勢"海峽兩岸圖書館事業研討會論文集，1997 年 5 月 25-28 日，頁 163-174.
- 註 4：陳昭珍，"電子圖書館資訊組織問題之探討，"海峽兩岸圖書館事業研討會論文集，1997 年 5 月 25-28 日，頁 175-196。
- 註 5：Lorcan Dempsey, "ROADS to Desire," <http://hosted.ukoln.ac.uk/mirrored/lis-journals/dlib/dlib/july96/07dempsey.html>
- 註 6：Renato Iannella and Andrew Waugh, "Metadata: Enabling the Internet," <http://www.dstc.edu.au/RDU/reports/CAUSE97/>
- 註 7：Pamela Drew and Jerry Ying, "A Metadata Architecture for Multi-System Interoperability," <http://www.nml.org/resources/misc/metadata/proceedings/drew/metaarch.html>
- 註 8：Jeff Rothenberg, "Metadata to Support Data Quality and Longevity," http://www.nml.org/resources/misc/metadata/rothenberg_paper/ieee.data-quality.html
- 註 9：梁中平、徐千惠，「取 SGML 之長，補 HTML 之短—新一代標示語言 XML」
（民 86 年 11 月），(<http://www.cals.org.tw/files/cals1-4.htm>)
- 註 10：Arbortext, "XML for Managers",
(http://www.arbortext.com/Think_Tank/XML_Resources/xml_for_managers.html)
- 註 11：Dave Raggett, Arnaud Le Hors and Ian Jacobs, ed. "HTML 4.0 Specification", [W3C Recommendation], (Apr 1998),

(<http://www.w3.org/TR/1998/REC-html40-19980424>) [7.5.5]

註 12：Jon Bosak, “Media-Independent Publishing : Four Myths about XML”, IEEE Computer Vol.31, No.10 (Oct 1998), p.120-122

(<http://metalab.unc.edu/pub/sun-info/xml/why/4myths.htm>)

註 13：Jon Bosak, “XML, Java, and the future of the Web” (Sun Microsystems, Mar 1997) (<http://sunsite.unc.edu/pub/sun-info/standards/xml/why/xmlapps.htm>)

註 14：Ralf I. Pfeiffer, “XML Tutorial for Programmers : Writing XML Documents”, (IBM XML Technology Group),

(<http://www.software.ibm.com/xml/education/tutorial-prog/writing.html>)

註 15：同註 14，[9.1]。

註 16：張錦堂，”Chinese XML FAQ” (中央研究院計算中心，Dec 1998)

(<http://www.ascc.net/xml/zh/big5/faq.html>)

註 17：同註 14，[5.3.2]。

註 18：Metalogy 系統主要發展人員包括：陳昭珍、陳雪華、陳光華、余顯強、陳嵩榮、蔡理暉、謝順宏、歐陽慧。

參考文獻

- A fine metadata mess is in prospect. *Digital Publishing Technologies* 2:7(Jul 1997):6.
- IT Link. *Organising the Internet* 9 (11) (Dec. 1996):6-8.
- OCLC awards 1996 research grants. *OCLC Newsletter* 222(Jul/Aug 1996):22.
- Aalberg, T."Uniform access to collections with visual information: main subject project at NTNU, Institute for Data Technology and Information Science". *Synopsis* 28:2(Apr 1997):75-80.
- Allen, R. S. "An overview of the Federal Geographic Data Committee National Spatial Data Infrastructure, National Geospatial Data Clearinghouse, and the Digital Geospatial Metadata Standard: what will it mean for tomorrow's libraries"?. *Geography and Map Division Bulletin* 180 (Jun 1995):2-28.
- Andersen, J. "Internet and registration: a commentary". *DF Revy* 20:3(Apr 1997):75-6.
- Andresen, L."Metadata: new key concept in Internet circles". *Bibliotekspressen* 6(Mar 1997):152-3.
- Armstrong, C. J." Metadata, PICS and quality". *Online and CD-ROM Review* 21:4(Aug 1997): 217-22.
- Ashrafi, N & Kuilboer, J-P. "The information repository: a tool for metadata management". *Journal of Database Management* 6:2(Spring 1995):3-12.
- Bearman, D. & Duff, W. "Grounding archival description in the functional requirements for evidence". *Archivaria* 41 (Spring 1997):275-303.
- Bearman, D. "CNI/OCLC Workshop on metadata for networked images". *Archives and Museum Informatics* 10:3(1996):270-6.
- Bearman, D. "Developments in metadata management frameworks". *Archives and Museum Informatics* 10:2(1996):185-8.
- Bearman, D. "How the information revolution might affect us professionally".

- Canadian Journal of Information and Library Science 22:1(Apr 1997):38-55.
- Bearman, D. "Possible contributions of the reference model of metadata required for evidence to a reference model of metadata required for image description". Archives and Museum Informatics 10:3(1996):295-302.
- Bide, M. & Green, B. "Electronic tables of contents (EToCs) for serials: first steps towards standards for metadata". Serials 8:3(Nov 1995):257-60.
- Brodie, M. "In the company of strangers: challenges and opportunities in metadata implementation". LASIE 28:4(Dec 1997):40-6.
- Butenfield, B. P. "Geographic information systems and digital libraries: issues of size and scalability. Geographic information systems and libraries: patrons, maps and spatial information". Illinois University at Urbana-Champaign (1996):69-80.
- Campbell, D. G. & Cox J. P. "Cataloguing Internet resources". Feliciter 43:5(May 1997):60-3.
- Caplan, P. & Guenther, R. "Metadata for Internet resources: the Dublin Core Metadata Elements Set and its mapping to USMARC". Cataloging & Classification Quarterly 22: 3/4(1996):43-58.
- Chen, P-W. & Cha, e. S-K. "Knowledge-based multimedia information retrieval in hyperspacng". Telematics and Informatics 14:1(Feb 1997):27-50.
- Cheung, W. C & Hsu, C. "The model-assisted global query system for multiple databases in distributed enterprises". ACM Transactions on Information Systems 14:4(Oct 1996):421-70.
- Ciganik, M. "Metainformation as a tool for information system management". Kniznice a Informacie 28:11(1996):441-8.
- Clemson, P. A. "An inside approach to a networked document cataloging". Journal of Internet Cataloging. 1:2(1997):57-64.
- Cobb, D. & Olivero, A. "
- +The Massachusetts Electronic Atlas: an interactive Web site for access to maps and geographic data for the Commonwealth of Massachusetts". Journal of Academic Librarianship 23:3 (May 1997):231-5.

- Cox, R. J. "More than diplomatic: functional requirements for evidence in recordkeeping". *Records Management Journal* 7:1(Apr 1997):31-57.
- Crook, M. "Research Advisory Committee meets". *OCLC Newsletter* 217(Sep/Oct 1995):13-14.
- Damcevski, P.& Chatwin, D. "Facts about the figures: issues in visibility and accessibility of metadata in numerical datasets". *Cataloguing Australia* 21:3/4(Sep/Dec 1995):182-4.
- Dempsey, L. "Metadata". *LASER Link* (Spring/Summer 1997):16-17.
- Dempsey, L. "Metadata". *LASER Link* (Spring/Summer 1997):16-17.
- Dempsey, L. "Metadata: a UK HE perspective". *British Library. Research and Innovation Report* 78(1997):133-5.
- Desai, B. C. "Supporting discovery in virtual libraries". *Journal of the American Society for Information Science* 48:3(Mar 1997):190-204.
- Domaratz, M. "Finding and accessing spatial data in the National Spatial Data Infrastructure. Geographic information systems and libraries: patrons, maps and spatial information." *Illinois University at Urbana-Champaign*(1996):31-40
- Duff Will, W. "Metadata replace archival description: a commentary". *Archivaria* 39 (Spring 1995):33-8.
- Duff, W. "Ensuring the preservation of reliable evidence: a research project funded by the NHPRC". *Archivaria* 42 (Fall 96):28-45.
- Feldman, S. "Advances in digital libraries '97: the Washington IEEE forum delves into the implications of technology". *Information Today* 14:7(Jul/Aug 1997):12-13,55.
- Ferguson, P. "ELDIS: mapping information on development and the environment". *Information Development* 12:3(Sep 1996):155-62.
- Hakala, J.; Koch, T. & Husby, O. "Warwick framework and Dublin core set provide a comprehensive infrastructure for network resource description". *Nordinfo Nytt* 2(1996):40-8.

- Heery, R. "Review of metadata formats". Program 30:4(Oct 1996):345-73.
- Henze, V. & Schefczik, M. "Metadata. Connection between Dublin Core Set, Warwick Framework and data formats". Bibliotheksdienst 31:3(1997):413-419.
- Hurley, C. "Ambient functions: abandoned children to zoos". Archivaria 40 (Fall 1995):21-39.
- Jizba, L. "Reflections on summarizing and abstracting: implications for Internet Web documents, and standardized library cataloging databases". Journal of Internet Cataloging 1:2(1997):15-39.
- Kerr, M. "NewsAgent for Libraries". Vine 104 (1996):51-3.
- Kingwell, J. "Trends in interactive information systems for earth observation from space: Towards a global, digital image library service". Australian Special Libraries 29:2(Jun 1996):49-60.
- Lamb, P. C. "Seven products you can sell online". Digital Publishing Strategies 1: 12(Aug 1997):9.
- Lehmann, K-D. "European national libraries and the CoBRA Forum of the EU Libraries Programme". Alexandria 8:3(1996):155-66.
- Lehmann, K. D. "Difficulties with the levels. Cataloguing rules, data formats and communication interfaces". Zeitschrift für Bibliothekswesen und Bibliographie 44:3 (May/Jun 1997):229-40.
- Lester, R. "Internet hype and Web wonders". Assistant Librarian 89:7(Jul/Aug 1996):109-11.
- Lichtenberg, J. "Inching towards e-commerce". Publishers Weekly 244:50 (Dec 1997):35-7.
- Lunin, L. F. "ASIS, FID hold celebratory meeting in Chicago". Information Today 12 :11 (Dec 1995):25-6.
- Lynch, C. "A summary of the findings of CNL's NIDR research initiative. British Library." Research and Development Department. BLRD Report 6250(1996):89-93.

- Lynch, C. "Metadata: moving from planning to implementation". British Library. Research and Innovation Report 78(1997):137-9.
- MacNeil, H. "Metadata strategies and archival descriptions: comparing apples to oranges". *Archivaria* 39 (Spring 1995):22-32.
- Mangan, E. U. "The making of a standard". *Information Technology and Libraries* 14:2(Jun 1995):99-110.
- Marsh, E. C. etc. "Looking ahead: the OCLC Office of Research". *OCLC Newsletter* 225 (Jan/Feb 1997):14-26.
- Neville, L. "Internet resources: issues of access and possible solution". *Cataloguing Australia* 23:1/2(Mar/Jun 1997):23-7.
- Nordinfo, S. v. & Nytt. "Ungern-Sternberg Information day on metadata in Lund". *Informationsdag om metadata i Lund* 3 (1996):13-16.
- Paskin, N. "Information identifiers". *Learned Publishing* 10:2 (Apr 1997):135-56.
- Perkins, M. "Why don't search engines work better?." *Information Outlook* 1:5(May 1997):40-1.
- Phillips, J. T. "Metadata - information about electronic records". *Records Management Quarterly* 29:4(Oct 1995):52-5.
- Pountain, D. "Of teams and components". *Byte* 22:5(May 1997):40IS15-16, 40IS18, 40IS20,40IS22.
- Queau, P. "Meta-data". *IFLA Journal* 23:1(1997):20-2.
- Rand, R. Y. etc. "Libraries, global change data, and information management". *Library Hi Tech* 13:1-2 (1995):26-42.
- Rochester, M. "Report on the 63rd IFLA conference, Copenhagen, August 31-September 5, 1997". *Education for Library and Information Services: Australia* 14:3(Nov 1997):62-4.
- Rogers, D. "Cataloguing Internet resources: the evolution of the Dublin Core metadata

- set". *Cataloguing Australia* 23:1/2(Mar/Jun 1997):17-22.
- Rotter, W. MALVINE. "A digital project of the Austrian Literature Archive".
Mitteilungen der Vereinigung Österreichischer Bibliothekarinnen und Bibliothekare 50:2(1997):70-2.
- Rusch-Feja, D. "Subject-oriented collection of information resources from the Internet". *Libri* 47:1(Apr 1997):1-24.
- Schram, S. G. & Wallace, W. W. "USDA Global Environmental Change Data Assessment and Integration Project". *Journal of Agricultural & Food Information* 3:2(1995):39-47.
- Shoklar, K.; Basu, C. & Shah, K. "Putting legacy data on the Web: a repository definition language". *Computer Networks and ISDN Systems* 27:6(Apr 1995):939-51.
- Simth, J. T. "Meta-analysis: the librarian as a member of an interdisciplinary research team". *Library Trends* 45:2(Fall 1996):265-79.
- Smits, J. "Digital metadata, standards for communication and preservation". *European Research Libraries Cooperation*. 6:4(1996):383-406.
- Sorensen, A. "Registration in the future: new demands, new media; keywords from the DF conference on 4 and 5 February 1997". *DF Revy* 20:3(Apr 1997):71-4.
- Sprehe, J. T. "The U.S. Census Bureau's data access and dissemination system (DADS)". *Government Information Quarterly* 14:1(1997):91-100.
- Steele, C. "Organization of knowledge on the Internet". *Information Management Report* (Sep 1996):1-6.
- Stratford, J. S. & Stratford, J. "Computerized and networked government information: California Environmental Resources Assessment System". *Journal of Government Information* 22:3(May/Jun 1995):267-72.
- Tammaro, A. M. "Cataloguing, cataloguing ... metacataloguing". [Catalogando, catalogando ...metacatalogando.] *Biblioteche Oggi* 15:1(Jan-Feb 1997):80-14.
- Tomimoto, H. "Systematization of network resources". *Igaku Toshokan (Journal of*

- the Japan Medical Library Association) 44:1(Mar 1997):54-65.
- Vidgen, R. "Temporal object modeling: diagramming conventions and design considerations". *Journal of Database Management* 8:1(Winter 1997):14-23.
- Vidgen, R. "Temporal object modelling: diagramming conventions and design considerations". *Journal of Database Management* 8:1(Winter 1997):14-23.
- Vizine-Goetz, D. "OCLC investigates using classification tools to organize Internet data". *OCLC Newsletter* 226 (Mar/Apr 1997):14-18.
- Wallace, D. A. "Managing the present: metadata as archival description". *Archivaria* 39(Spring 1995):11-21.
- Watson, B. C. "Research Advisory Committee meets, gets update on OCLC Office of Research projects". *OCLC Newsletter* 229(Sep/Oct 1997):16-18.
- Watson, B. C. "Research Advisory Committee meets". *OCLC Newsletter* 226 (Mar/Apr 1997):19-20.
- Watson, B. "Research Advisory Committee meets". *OCLC Newsletter* 223(Sep/Oct 1996): 18.
- Weibel, S. & Miller, E. "Cataloging syntax and public policy meet in PICS". *OCLC Newsletter* 227(May/Jun 1997):28-9.
- Werf-Davelaar, T. v. d. "The bibliographical description of electronic information resources. 2. *Informatie Professional* 1:7/8(1997):25-31.
- Wheatley, A. & Armstrong, C. J. "Metadata, recall, and abstracts: can abstracts ever be reliable indicators of document value?." *Aslib Proceedings* 49:8(Sep 1997):206-13.
- White, M. "Intranets aren't always the solution". *Digital Publishing Strategies* 2:3(Nov 1997):6.
- Wu, C-J. "Comparison of three metadata standards". *Bulletin of the Library Association of China* 57(Dec 1996):35-45.
- Younger, J. A. "Resources description in the digital age". *Library Trends* 45:3(Winter 1997):462-87.

附件：

出席國際學術會議報告

會議名稱： 博物館資訊交換協會會議

CIMI (Consortium for the Computer Interchange
of Museum Information) Meeting

會議時間： 2001 年 6 月 21 日～23 日

會議地點： 美國紐約

行政院國家科學委員會補助國內專家學者出席國際學術會議報告

90 年 7 月 31 日

報告人姓名	陳光華	服務機關及職稱	國立臺灣大學圖書資訊學系 副教授
會議時間 地點	2001/06/21-2001/06/23 美國紐約 (New York)	本會核定 補助文號	NSC 89-2750-P-002-014-
會議名稱	(中文)博物館資訊交換協會會議 (英文)CIMI (Consortium for the Computer Interchange of Museum Information) Meeting		
發表論文 題目	無		

報告內容目次

一、CIMI 簡介	03
二、參加會議經過	05
三、議程	06
四、與會人員	08
五、與會心得	10
六、建議	11
七、會議結論	12
八、附件	13
1. Sample catalogue record fields #1	13
2. Sample Catalogue Record Fields #2	14
3. Sample Location History & Audit Fields	14

一、 CIMI 簡介

(一) 簡介

CIMI (Consortium for the Computer Interchange of Museum Information)成立的宗旨在於：將博物館資訊普及社會大眾，推展開放式的系統標準，以管理及傳遞數位博物館資訊。自 1990 年成立以來，CIMI 在博物館電子資料架構的研究上，已有相當的進展，使其搜尋檢索功能更為廣泛。主要工作係透過合作的實驗計畫，研擬如何將資料標準化，以利電子方式查詢。

(二) 宗旨

CIMI 會員機構透過下列方式，致力推展「文化遺產」資訊之普及：

1. 推動開放式系統標準，以記錄電子文化遺產資訊、分享資源。
2. 在實驗計畫中，以標準系統處理博物館資訊；並邀請會員參與，進而幫助會員們發展自己的資訊體系。
3. 在協會中分享這些合作成果。
4. 彼此交換有關電子資訊議題的心得。

(三) 工作內容

CIMI 目前的五大工作重點，旨在建立數位博物館資訊的普及性，並發展健全的管理體系，永續經營。

1. 整合資訊管理

CIMI 目前正在研擬如何就不同機構的資源層次，提供不同的使用工具與標準，以使各機構能依其需要、按其預算，發展他們自己資訊管理的整合體系。

2. 資源發現與探索

CIMI 早期工作的重點大多在建立資訊架構的標準，以利使用者以電子方式取得。目前 CIMI 方針已邁入研擬發現探索資訊的標準及方法。

3. 分散式的搜尋與檢索

CIMI 目前進行測試的方式，在 Internet 之搜尋檢索功能將不亞於一個 in-house 系統。

4. 標準化的磋商

標準若要成為真正的標準，必須具有強烈的共識。透過仔細的協商與嚴謹的實

驗，致力推動「文化遺產」相關單位對電子資訊標準的共識。

5. 配合國際發展

全球有許多團體也在研發資訊交換的標準。CIMI 與這些團體保持連繫，並致力配合國際的相關發展。

二、參加會議經過

CIMI 乃自 MCN (Museum Computer Network) 分離出來的組織，成立的目的主要在探討電子博物館相關標準，初期的會員雖然不多，但隨著近幾年的努力，並致力於博物館各項標準的制訂與實驗，建立了國際性的聲望，會員也陸續增加。CIMI 的工作人員不多，但卻推動了很多計畫，如 CHIO (Cultural Heritage Information Online)、Z39.50、Dublin Core、IIM (Integrated Information Management)、SPECTRUM XML、OAI、DOI 等，主要乃以 Testbed (實驗計畫) 的方式在運作。

CIMI 每年舉辦四次工作會議，亦即每季一次，分別於三月、六月、九月、十二月舉行，本次會議為 CIMI 六月的夏季工作會議。這次會議的主要議題為討論 Harmony 實驗計畫、XML 實驗計畫，並針對 OAI 以及 DOI 等各項標準在未來圖書館／博物館資訊交換所扮演的角色，進行討論。

三、議程

本次會議為本年度第二次會議，於美國紐約大都會美術館與古根漢美術館舉行。議程為期三天（六月二十一日到二十三日），其中六月二十一日為會前會在美國大都會美術館舉辦 Dublin Core Forum，後兩天則是 CIMI 正式會議在古根漢美術館召開。詳細議程如下：

第 0 天：2001-06-21 (Thursday)

Dublin Core Forum (Stu Weibel, Director of DCMI)

第一天：2001-06-22 (Friday)

上午

1.1 Introductions and Meeting Overview

1.2 Project Updates

- Harmony Collaboration (Carl Lagoze)
- Open Archives Initiative (John, Henry, Carl)
 - outcomes of alpha test
 - strategy for future metadata harvesting activities

1.3 Guest speakers: Stephen Mooney, Director of New Business Development, International Digital Object Identifier (DOI) Foundation; Hal Espo and David Sidman, Content Directions

- Overview of the IDF and DOI. DOI demonstration. Followed by discussion of possible uses within the museum community.

下午

1.4 XML Testbed

- define strategy for future metadata harvesting activities
- begin discussing possible uses for DOI's within the museum community
- discuss work to date on SPECTRUM XML testbed

5:30-7:30pm Reception hosted by the Solomon R. Guggenheim Museum

第二天：2001-06-23 (Saturday)

上午

1.5 XML Testbed Update

下午

1.6 Project Updates

- ☐ Handscape (Cary)

1.7 CIMI Institute

- ☐ MCN 2001 conference program
- ☐ grant proposal
- ☐ ICHIM 2001 workshop

1.8 New Business

- ☐ CIMI Programming 2002

1.9 Membership Issues

- ☐ discuss current status of Handscape project
- ☐ discuss program and plans for MCN conference
- ☐ discuss possible work items for 2002

四、與會人員

參加本次會議的人員來自許多國家，如英國、美國、加拿大、荷蘭、澳洲、以及台灣，共計 37 位。

姓 名	單 位
Lynn Underwood	Solomon R. Guggenheim Museum
John Perkins	CIMI Executive Director
Angela Spinazze	Team CIMI Programs Manager
Alice Grant	CIMI Project Manager
Henry Stern	Team CIMI OAI Protocol Implementor
Martha Dunlop	Team CIMI Administrative Support
Barry Boyce	Team CIMI Writer and Editor
Scott MacCallum	CHIN
Liz Appleby	Willoughby Associates
Andrea Boyes	Willoughby Associates
Lenore Sarasan	Willoughby Associates
Bert Degenhart Drenth	ADLIB Information Systems
Paula Willey	Gallery Systems
Peter Manuelian	Museum of Fine Arts, Boston
Basil Dewhurst	Australian Museums On-Line
Tony Gill	Research Libraries Group
Pete Johnston	UKOLN
Neil Thomson	Natural History Museum, London
Len Steinbach	Cleveland Museum of Art
Jay Hoffman	Gallery Systems
Sherry, Hsueh-hua Chen	National Taiwan University
Kevin, Kuang-hua Chen	National Taiwan University
Joyce, Chao-chen Chen	National Taiwan Normal University
Theresa Slowik	Smithsonian American Art Museum
Mary Webb	AskArt
Susan Chun	Metropolitan Museum of Art
John Aubry	American Museum of Natural History
Rob Bruno	American Museum of Natural History Library

姓 名	單 位
Steve Mooney	International DOI Foundation
Hal Espo	Content Directions
David Sidman	Content Directions
Jay Hoffman	Gallery Systems
Danny Scheman	Mediafarm (representing Museum of Americas Foundation)
Gordon McKenna	mda
Michael Selway	System Simulation Ltd
Carl Goodman	Museum of the Moving Image
Larry Kooper	Metropolitan Museum of Art

五、與會心得

本次參加會議之心得主要如下所示：

1. 博物館領域的 METADATA 需求

Dublin Core 已經為國際性標準，但這次在會議中也發現博物館在利用這樣的標準的同時，有許多特殊的需求必須考慮，因此將 DC 視作是所謂的 Crosswork，也就是交換格式。XML 作為各種 metadata 描述語言的地位已經確立，英國的博物館相關機構，正推動 SPECTRUM，而 CIMI 也進行 SPECTRUM XML 實驗計畫。國內的博物館、美術館也應該思考這些問題，進行博物館資訊的數位化與標準化。

2. 系統廠商的參與

此次會議中有不少的系統廠商參與，其直接的影響就是藉由參與標準制訂的討論，在設計系統時能夠更清楚精確的掌握資料記錄標準上的要求，以及各不同類型館藏單位在處理資料上的特有需求，以更具彈性的方式設計系統功能。同時對 CIMI 這樣以虛擬方式運作的聯盟，可經由廠商的參與取得技術上的支援，以及立即測試標準的可行性與有效性。

3. 準標準的評估、實驗、與推動

CIMI 推動的實驗性計畫的特色，是對於可能成為標準的各項技術進行評估，若是覺得可行性甚高，便推動實驗性計畫，如果結論是肯定的、正面的，便會透過會員所屬的機構、相關的會議，推廣這些技術。

六、建議

1. 由於 CIMI 在博物館界的角色越來越重要，參與會員多數是博物館、美術館，我們則屬於學術性研究計畫團體的身份與會，我們建議國內重要的博物館、美術館應積極參與 CIMI。
2. 目前國內研究經費越來越緊縮，而參與 CIMI 的會費需 5000 美元，若以計畫型式編列該筆經費並不適合。我們認為若是國內博物館與美術館能夠編列預算，積極參與 CIMI，應是比較好的作法。
3. 積極研究電子博物館之相關標準，使博物館系統具有互通性及國際性，邀集國內各單位參加實驗計畫。
4. SPECTRUM XML DTD 是由英國博物館界推動的標準，目前也廣受各國重視，CIMI 亦進行 SPECTRUM XML DTD 實驗計畫，值得我國博物館界參考。
5. DC 已成為國際的標準，應推動國內從事 Metadata 研究的機構，將 DC 視為是各種不同類型 metadata 的 Crosswork。
6. DOI 以及 OAI 等各項技術或標準也廣受重視，國內教育部以及文建會等單位，亦陸續執行相關的計畫，評估或研究這些標準，CIMI 相關計畫的進展也可做為國內的借鏡。
7. CIMI 工作會議的召開地點目前仍以北美與歐洲為主，而亞洲地區的參與國僅有台灣，因此我們更有機會扮演亞太地區發言人的角色，可以將不同類型的需求，反應在各項標準的制訂過程中，使得國際標準符合我們的需求。
8. 由於 CIMI 是以工作計畫的方式維持每年的運作，因而每年度參與的個人及單位不一定相同，造成各年度 CIMI 運作情形有好有壞。然而，CIMI 確實是一個博物館界交流與討論的工作典範，如何在我國推動類似的機構，並排除前述的問題，是我國目前推動數位博物館計畫時，必須思考的課題。

七、會議結論

1. General version of DTD, version 0.9 must be frozen before testbed can proceed.
2. Current strict versions (0.1) need review before proceeding. Testbed group to do this via discussion list.
3. Process and decisions must be documented.
 - Determine method for documentation within testbed.
 - Determine method for formally sending mda feedback from testbed.
 - mda to articulate when they will go back to generalized DTD to update it with a new version number.
 - Versioning and process control must be strict and rigorous. (See OAI website for ideas.)
 - Define process for contributing data.
 - What format?
 - How many records?
 - Decision: Where there is a local association with a technical partner; the museum will send an export to the technical partner who will process it and send the resulting file to Alice.
 - Decision: Where there is a museum contributing content and acting as technical partner (Minnesota Historical, National Museum of Denmark) data must be delivered according to guidelines (to be determined by Alice -- bottom line: the file that is delivered will be in XML and conform to the strict version of the DTD along with all other "fall out" data).
4. Request content for person and object only, from appropriate testbed participants.
5. Map what is delivered against the strict DTDs (v0.1) for person and object. (both technical members and content providers)
6. Reconcile "fall out" with strict DTDs--what recommendations will be made to mda as a result of the data that doesn't fit?
7. How to capture (i.e., document) gaps between what is included in strict DTD and actual data contributed?
 - Suggested solution: each technical partner defines a namespace with all declarations of additional elements; top of XML document is place to declare which namespaces are in use; in contributed XML content = declaration in the XML document that points to namespace for the strict schema and points to the namespace for the contributor specific schema that describes the fall out (Bert will write instructions)

八、附件

Sample catalogue record fields #1

Object Name
 Collection
 Created by
 Created on (date)
 Amended by
 Amended on (date)
 Descriptive Line
 Title
 Text
 Type
 Attribution
 Name
 Role
 Uncertainty
 Attribution
 Date
 Role
 Attribution
 Place
 Role
 Uncertainty
 Physical Description
 Materials & Techniques
 Condition

Code
 Curator
 Date
 No. of Parts
 Member of ...
 Group
 Members
 Dimensions
 Type
 Value
 Units
 Qualification
 Subject
 Object History Note
 V&A Exh' History
 Cat' No./Date
 Title/Venue
 Textural References
 Citation
 Environmental Requirements
 Code
 Humidity
 Light
 Pollution
 UV:
 Temperature

Sample Catalogue Record Fields #2

%N mus_cat_M
 %F (sortexpr bm_object_number_expr)
 %F mus_other_number_type
 %F mus_other_number
 %F bm_bib_xref
 %F bm_bib_spec
 %F bm_bib_spec
 %F bm_coll_name
 %F bm_acq_year_text
 %F
 mus_object_production_person_association
 %F mus_object_production_person
 %F
 mus_object_production_place_association
 %F mus_object_production_place
 %F mus_object_production_date_text
 %F mus_object_production_date_period
 %F mus_inscription_type
 %F mus_inscription_content
 %F mus_inscription_language
 %F mus_inscription_translation
 %F mus_creation_date
 %N mus_cat_j

Sample Location History & Audit Fields

Museum Number
 Collection
 Object Name
 Acquisition
 Created by
 Created on (date)
 Amended by
 Amended on (date)
 Descriptive Line
 Classification
 Location
 Site
 Room
 Case
 Shelf
 Box
 Action By
 Date
 Current location
 Location
 Moved by
 Moved on (date)
 Location history
 Location
 Moved by
 Moved on (date)
 Location history
 Location
 Moved by
 Moved on (date)
 Loan status
 Borrower