

# 臺灣地區中國古籍數位化的現況與展望

第三次兩岸古籍整理研究學術討論會·國家圖書館舉辦

2001年4月18-19日

羅鳳珠·元智大學中語系

## 壹、前言

臺灣地區中國古籍數位化的發展，自中央研究院1984年7月開始的「史籍自動化計畫」迄今（2001年）已有十七年歷史。十七年來，參與的單位從中央研究院到故宮博物院、國家圖書館、歷史博物館、台灣省文獻處、文建會……等政府單位，大學院校也陸續投入。

在發展的過程中，中研院院內的「史籍自動化計畫」到各單位各自成立的研究群組與研究計畫，都是單位內自發性的研究，1998年起，國科會率先有計畫的規劃大型的「數位博物館專案先導計畫」、2000年行政院科技顧問組委託中研院規劃「國家典藏數位化計畫」、2001年文建會規劃「全國文化資料庫」計畫，都是由政府相關單位邀請學者有計畫的規劃、推動、執行。數位化的典藏範圍，也從文字資料擴充到圖片、文物、聲音等多媒體資料。在功能方面，除了原有的典藏、流通、研究、教學等功能，更進一步爲了資料的交換、典藏之需要，設計中文後設資料（Metadata，或譯爲詮釋資料）的格式，並與XML（eXtensible Markup Language）技術結合，以協助使用者更方便、準確的從數位資料庫中擷取資料，未來如能就資料內容建立內容標誌的標準，全面進行內容標誌

（Content Markup）的工作，將可使數位資料庫的使用邁向一個新的境界。在人才培訓方面，從早期仰賴資訊人才，培養研究助理協助計畫執行，迄今已有大學成立系所培養相關專長的學生，也有不定期舉辦的研習班提供大眾研習。在推展科技與人文結合的理念方面，前後有「科技與人文對話網路論壇」與「數位世紀的高峰會談」等論壇在公眾媒體呈現。在應用於教學方面，教育部卓越計畫、國家數位博物館計畫、教育部提升大學基礎教育計畫引用文獻數位化的成果作爲教學輔助教材。

## 貳、發展的過程與方式

### 一、發展的過程

筆者於〈台灣地區中國古籍文獻資料數位化的過程與未來的發展方向〉（註一）一文述及中國古籍文獻資料數位化的過程，分爲：「第一期：處理中文文字資料時期；第二期：單機版古籍全文資料庫的研發；第三期：網路版古籍全文資料庫的研發；第四期：多功能、多媒體、多元化的文獻資料庫；第五期：以三D（dimension）動畫技術呈現立體文獻資料。」五個階段。這五個階段轉型的契機，固然隨著資訊技術的進步而改變，參與人員的背景與參與單位投入的方式、整體環境的配合，對於發展的速度與方向也有舉足輕重的影響。

## 二、發展的方式

臺灣地區中國古籍數位化發展的方式，可從「主導與規劃者的學門」、「執行的單位」、「文獻的內容」、「資料庫的功能」、「推廣的方式」等方面來討論。

### (一) 主導與規劃者的學門：

可分為「由資訊學門規劃、主導」、「資訊學門與人文學門共同規劃」、「由人文學門主導」、「由文獻典藏機構主導」四個角度。四個角度的發展，各自受到不同環境背景因素影響而形成，並無主從上下的關係。

#### 1. 由資訊學門規劃、主導

中文數位化發展之初，需先解決中文的編碼與輸入、輸出的問題。處理中文文字資料的時期，以發展中文資訊技術及中文字形為主，中文文字的檢字法與編碼的研發工作為日後的中國文獻數位化工作奠定了基礎，是中國文獻數位化的萌芽階段，這一個階段結合了研究單位、學術單位、產業單位三方面努力的成果。

單機版古籍全文資料庫的研發，由中央研究院「史籍自動化計畫」小組負責，結合中研院資訊所以及史語所共同研發建置，完成《史記》、《漢書》、《後漢書》、《三國志》前四史，這個階段由資訊學門規劃、主導，邀請人文領域共同參與。由前四史擴充到二十五史，古籍全文資料庫的技術已臻於成熟。

#### 2. 資訊學門與人文學門共同規劃

中研院在完成二十五史，古籍全文資料庫的技術成熟之後，開始嘗試各種文獻資料庫的開發。有 1985 年 10 月開始的「漢代墓葬綜合研究資料庫」，1986 年 2 月的「台灣土著語言資料庫」，1986 年 4 月的「台灣日據時代戶籍資料庫」，1987 年 1 月的「清代竹塹地區土地申告書資料庫」，以及 1989 年計算中心所做的「說文解字和玉篇資料庫」等等。利用影片處理技術所做的古籍資料庫，如傅斯年圖書館發展的「善本書影片資料庫」，此時處理的技術已經從全文資料庫技術發展到關聯式資料庫。參與的單位，也從資訊所、史語所增加到臺史所、近史所、文哲所、圖書館，技術全由中研院自行開發，此時全文資料庫的製作技術已趨成熟，1995 年開始，中研院陸續與國內外建立發展共享古籍資料的合作關係。（註二）

#### 3. 由人文學門主導

除中研院之外，東吳大學陳郁夫教授也陸續推出單機版的《十三經》、《宋儒學案》、《明儒學案》等全文檢索資料庫；這些資料庫均以文字資料為主。筆者於 1993 年完成《全唐詩》、《紅樓夢》多媒體全文檢索資料庫，進一步結合文字、影片、聲音、圖片資料，嘗試多媒體資料庫的研發工作。中研院文哲所的《道藏》資料庫、歷史所的古文字資料庫等，由於主持計畫者都是人文背景，資料庫的建置方

式，由人文學門規劃，除了陳郁夫教授能親自撰寫程式，其餘均由資訊學門協助技術的開發。

#### 4. 由文獻典藏機構主導

文獻典藏機構收藏豐富的文獻資料，扮演文獻典藏與流通的角色，資訊發展之後，將文獻資料數位化，除了使文獻資料以不同形式的媒體保存之外，數位化的文獻資料，可提供使用者更方便的流通管道。國家圖書館明人詩文集、故宮博物院的書畫文物、中央研究院傅斯年圖書館善本書影片資料庫、中央研究院歷史所文物圖像資料庫、台灣大學淡水河溯源計畫的淡新檔案、伊能文庫、岸裡大社文書、人類學系所藏古文書等，都是由典藏單位主導文獻數位化的工作。

#### (二) 執行的單位

台灣地區文獻數位化工作，最初是由研究單位自發性的於單位內部提出研究計畫，邀請理念相同的人共襄盛舉，由執行單位視需要與其他單位合作，如中研院的「史籍自動化計畫」；而後由研究人員向相關單位提出研究計畫，申請研究經費，如國科會研究計畫；或以個人的力量集合志同道合的人共同完成。近年來已由個人及單位內部的單打獨鬥，發展為由單位有計畫的發展整合型計畫，如中研院文哲所漢「學研究資料庫研發計畫」、中研院語言所「語言典藏與語言座標計畫」、台灣大學與中華佛學研究所「電子佛典數位化計畫」。1998年由國科會主導的「數位博物館專案先導計畫」開始，政府部門有計畫的規劃文獻數位化的工作，如今已進入第三年；1999年7月行政院「電子、通訊、資訊策略會議」通過「國家數位典藏計畫」，第一年的先導計畫已自2001年1月份開始執行，2001年文建會數位文化網路發展計畫的「全國文化資料庫計畫」也將在今年度啟動。至此為止，台灣地區的文獻數位化工作，已由政府單位結合學術單位、典藏單位、產業界、民間力量共同執行。

#### (三) 文獻的內容：

從平面的圖書到立體的文物，從無聲的語文到有聲的語音，從紙本文獻到多媒體影音文獻，從靜態藝文圖片到動態影劇影片，從原典文獻到專書期刊等研究資料，古籍文獻數位化的資料包含早期文獻原典的文字、圖片、文物資料，後期聲光影像工業發展之後所記錄的聲音、影片資料，以及研究資料等。文獻數位化的內容隨著資訊技術的進步而從文字媒體拓展到多媒體，也隨著使用者的需要而從文獻原典資料擴充到研究資料。

古籍數位化發展之初，文獻內容以文字為主，其後增加表格、平面圖片（書、畫、拓片、書影……）、聲音（音樂、語音……）、影片、立體文物圖片、地理資訊資料。發展的關鍵全視資訊技術的開發而定。「史籍自動化計畫」1985年7月開始前四史資料庫的建置，迄1990年6月完成二十五史，內容以文字為主，不含表格，直

至 1995 年 3 月開始補充表格部份，至 1997 年 1 月完成。1993 年筆者所研發的單機版紅樓夢多媒體系統增加了圖片、聲音與影片資料。中央研究院歷史所文物圖像資料庫所建置的「簡帛金石資料庫」、「居延漢簡補編圖象檢索系統」、「武氏祠畫象檢索系統」、「安丘董家莊漢墓畫象」、「番社采風圖」等，都含有大量的圖片資料。故宮博物院的故宮文物之美系列，內含「明清琺瑯工藝」、「宋代書畫冊頁之美」、「佛經圖繪詳說」等，將故宮典藏的文物資料數位化，文獻內容包含平面書畫圖片及立體工藝文物。文建會網路劇院建置影片、聲音資料典藏網站，內容包含現代戲劇、傳統戲曲、舞蹈和音樂四類團體的資料。

在研究資料庫方面，以國家圖書館為首的「全國圖書聯合目錄」、「善本古籍聯合目錄」、「中文期刊聯合目錄」、「全國博碩士論文摘要檢索系統」等，中央研究院歷史所的「內閣大庫檔案目錄」、「唐宋明清史書目」，中央研究院文哲所漢學研究資料庫，筆者與國家圖書館、華盛頓大學東亞圖書館共同建置的「唐宋文學研究論著資料」等，都是以研究資料作為數位化的典藏範圍。

#### （四）資料庫的功能

這裡所要探討的資料庫功能，不是指單一資料庫所提供的功能，而是從整體的角度來探討從「史籍自動化計畫」以來，經過十七年的演變，資料庫大架構功能的發展情況。

1. 單機版單一資料庫的全文檢索功能：無論是純文字的《二十五史》、《十三經》等資料庫，或是多媒體的《紅樓夢》資料庫未改為網路版之前，都是以單機版的形式呈現，提供單一資料庫的全文檢索功能。
2. 網路版單一資料庫的文字資料全文檢索功能：早期開發的單機版全文檢索資料庫，在轉為網路版之後，依然提供單一資料庫的全文檢索功能。
3. 網路版多元資料庫的全文檢索功能：從建立學科資料中心的觀點，結合原典文獻以及研究資料的多元資料庫，如筆者所建置的「紅樓夢多媒體網路資料中心」。
4. 網路版跨資料庫的全文檢索功能：從某一個資料庫的查詢延伸到可以用相同的查詢條件跨到另一個資料庫查詢，如中央研究院地理資訊系統，在查詢二十五史的地名時，可跨入地理資訊系統之中國歷史地圖顯示其地理位置。
5. 跨資料庫的語文知識網路：將文獻資料整理之後，以不同的形式呈現，提供另一種新的知識結構，因應不同的使用需求，如「搜文解字語文知識網路系統」

#### （五）推廣的方式

古籍數位化是人文與資訊科技結合的起點，「史籍自動化計畫」發展之初，以中研院研究人員的專長與興趣為主要的考量，因此還不存在推廣的問題。當資訊網

路的風潮快速席捲各行各業各階層時，古籍數位化的發展便需要適時的推廣，以提升使用效益，也開始有更多人關注資訊科技與人文結合的方式，於是有了由行政院國科會主辦，聯合報協辦的「科技與人文對話網路論壇」之座談活動，除了邀請來賓現場座談，另外也透過網路擴大參與，主要目的在促進「科技」與「人文」二個領域人士的相互瞭解，減少科技發展對社會帶來的負面衝擊。2000年公共電視與中國時報也推出十三集的「數位世紀的高峰會談」節目，關注資訊時代的資訊政策及資訊素養的問題。當古籍數位化的功能從研究拓展到教學時，師資的培訓也成爲重要的一環，如國家數位博物館計畫定期舉辦「中小學教師數位博物館素養訓練班」，也是以教師爲對象的推展活動。文建會「全國文化資料庫計畫」將「舉辦系列推廣說明會」列爲計畫的配套發展計畫，其用意也在於推廣。

## 參、發展的現況與展望

台灣古籍數位化的工作，自1998年由國科會規劃的「數位博物館專案先導計畫」開始，大型的計畫已由政府單位推動，古籍數位化的工作已成爲有計畫、有規模的政策性工作。在資料整合方面，開始建立後設資料的標準格式，發展後設資料與XML技術結合，以提升資料檢索的完整性與精確性。

在古籍數位化專長人員的培育方面，從過去研究小組各自培養研究助理，到2000年，已有大學成立以培養中文與資訊結合人才爲主的科系成立。在功能方面，已從提供研究拓展到提供教學，使教學也能充分利用文獻數位化的成果。

### 一、大型計畫的推動

#### (一) 國科會數位博物館計畫：

1. 第一年的專案先導計畫，執行期間爲1998年12月1日至1999年10月31日，共有「不朽的殿堂－漢代的墓葬與文化」、「搜文解字－語文知識網路」、「臺灣原住民－平埔族群」、「淡水河溯源」、「蝴蝶生態面面觀」、「臺灣的本土植物和魚類」、「火器與明清戰爭」、「資源組織與檢索之規範」、「人文與自然資源地圖」、「數位典藏系統技術研發」、「系統評估－以「淡水河溯源」爲例」等十一個計畫，目前已完成上線。
2. 第二年計畫執行期間爲1999年12月1日至2000年10月31日，共有「故宮文物之美系列」、「玄奘西域行」、「淡水河溯源(II)」、「台灣民間藝術家之建置－以楊英風爲例」、「台灣建築史」、「生命科學－人體奧秘展覽館」、「台灣本土魚類(II)－尋回台灣本土的淡水魚類」、「語文知識網路(II)－生活語文、唐宋流行歌」、「平埔文化網路之維護與管理」、「中醫藥、針灸數位博物館」、「蘭嶼生物/生物多樣性數位博物館」、「台灣的老照片」等十二個計畫。

#### (二) 國家典藏數位化計畫

第一年計畫的執行期限自 2001 年 1 月 1 日至 2001 年 12 月 31 日，參與的單位與計畫包括：（註三）

1. 中央研究院：台灣原住民---平埔族數位化典藏計畫、台灣地方檔案數位化影像製作計畫、漢語及南島語語料庫計畫、近史所檔案館藏外交經濟重要檔案數位化典藏計畫、歷史語言研究所珍藏文物數位化典藏計畫、近代中國歷史地圖與航空照片資訊典藏計畫、台灣的動物相典藏之研究---魚類與貝類、台灣的本土植物數位典藏計畫、數位典藏技術支援與研發計畫。
2. 故宮博物院：故宮器物數位典藏子系統、故宮書畫數位典藏子系統、故宮清代檔案數位典藏子系統、故宮文物圖檔建置及管理子系統、故宮文物數位典藏電腦系統建置子系統。
3. 國立歷史博物館：建置歷史文物詮釋資料（Metadata）及索引典、建置智慧型歷史文物多媒體資料庫、建置歷史文物資料處理數位化制度。
4. 國家圖書館：國家圖書館古籍文獻典藏數位化計畫（明人詩文集、金石拓片）、台灣地區地方文獻典藏數位化計畫、國家圖書館期刊報紙典藏數位化計畫。
5. 自然科學博物館：動物學典藏數位化計畫、植物學典藏數位化計畫、地質學典藏數位化計畫、人類學典藏數位化計畫、數位典藏資訊技術整合計畫。
6. 台灣大學：台灣文獻文物典藏數位化計畫（淡新檔案全文及影像資料庫、岸裡大社全文及影像資料庫、伊能嘉矩手稿影像資料庫、日據時期台灣地區拓碑全文及影像資料庫、台灣舊照片資料庫）
7. 台灣省文獻委員會：台灣總督府檔案數位化子計畫、台灣總督府專賣局檔案數位化子計畫、台灣省行政長官公署檔案數位化子計畫。

### （三）文建會全國文化資料庫計畫：

2001 年，文建會在數位化文化網路發展計畫部分共編列二億元預算。其中預算高達九千八百萬元的「全國文化資料庫計畫」，是由台大資訊系教授項潔所規劃，日前剛完成期末審查。目前規劃的方向是要建立一個整合中央與地方文化資訊流通的數位化網路資料庫，同時，配套發展計畫包括管理資料庫人才的培訓、舉辦系列推廣說明會、訂定資料傳輸規定與管理機制。文建會表示，全國文化資料庫將援用 1996 年藝文資源調查所做的分類，將發展內容分為：信仰節俗、飲食文化、傳統聚落與傳統建築、人文景觀、考古遺址、西洋音樂、傳統音樂、傳統戲劇、現代戲劇、鄉土雜技、舞蹈、美術、文學、原住民文化、文化硬體展覽設施等十五項。（註四）

「國家文化資料庫」將整合各國所有文化相關機構各自成立的文化資料庫，構成全國文化資源的總目錄。在文建會的構想下，未來不僅所有公、私立博物館、美術館、收藏展示館、教育單位、研究機構中的典藏文物都將以數位化儲存，地方文

史工作者、藝術團體、創作者、業餘收藏家、文化藝術產業等，也都是資料提供者。「國家文化資料庫」將以文字、圖片、聲音、影片、地圖五種類型，收藏所有散布在台灣各地的文化資產、藝術創作，甚至人民村落的歷史。（註五）

## 二、資料整合的準備

文獻資料數位化所用到的技術非常多，典藏的文獻資料形式、內容、語言、媒體也很多樣化，參與數位典藏的機構也非常多，爲了不同資料庫之間的資料的交換、流通，爲了提升資料檢索的完整性與精確度，後設資料格式標準的建置以及與 XML 技術的結合，是發展文獻數位化的核心工作。國家圖書館於 2000 年開始，邀請國內相關的學者專家一起成立「Metadata 研究小組」，其目的有五項：「1、探討國內外 Metadata 發展現況、2、積極加入國內外 Metadata 研究組織及相關計畫、3、發展各主題領域或資料類型之 Metadata、4、研究 MARC 與 Metadata 轉換之語法、語意等相關標準、5、發展 Metadata 管理系統。」（註六）並且蒐集國內目前已經設計出來的 Metadata 格式匯集成《中文詮釋資料（Metadata）格式彙編》（同註六），以方便各單位取得及採用。參與國家典藏數位化計畫之國立歷史博物館，也有一個「建置歷史文物詮釋資料（Metadata）及索引典」子計畫，建置歷史文物詮釋資料格式，中研院歷史所第二年的「語言典藏的後設資料／語言識別」計畫，也將建立語言的後設資料。

## 三、人才培訓方面：

除了文建會全國文化資料庫計畫規劃管理資料庫人才培訓的配套措施之外，國內大學院校也開始注意到人文學門文獻數位化人才培訓的問題，台北大學中國語文學系是國內第一個以「期將中國語文予以資訊化處理，以培養優秀中國語文人才，使能系統化，科學化，便捷化處理及傳播語文資產；並應用資訊科技，從事當代語文研究，導引本土語文永續發展，開拓中文學術研究新領域，以促進國家社會進步。」（註七）爲設系宗旨的科系，其課程除了中國語文方面的課程之外，另有十門資訊類課程。東吳大學即將申請設立的古典文獻學研究所，預期也將朝著文獻與資訊整合的方向發展。大學院校之人文領域開始培養中文文獻數位化的人才，對於文獻數位化的長期發展有其正面的意義。

## 四、功能方面

過去的文獻數位化功能多數提供研究使用，國科會數位博物館計畫成立之宗旨，設定在建立文化、藝術、科技等教育性網站之內容典範，不僅要使典藏單位更了解文物數位化的重要性及規劃執行方式，也應使一般民眾體驗在網路上遨遊數位博物館的便利與樂趣，而網路教育觀念應從基層做起，因此「數位博物館專案計畫」必須推廣至中小學，中小學老師可有效利用網路建構其教學資料，使中小學生擁有正確使用網路的觀念。使用對象設定於中小學，因此計畫實施以來，每年都在北、中、南、東區舉辦中小學教師數位博物館素養訓練班，以培訓中小學師資之人文資訊素養。（註八）計畫亦舉辦教材設計比賽，其用意均在於將文獻數位化的功能，從典藏、流通、研究轉向教學。僑務委

員會所建立的華文教育中心也以教學為主（註九），2000年1月份開始執行的卓越計畫第一期計畫，其中清華大學主持的卓越計畫--「下一世代資訊通訊網路尖端技術及應用--網路教育園區及其社會影響研究」（註十），即以建置網路教學園區為主，將網路教學由學校擴充到社會。卓越計畫之第二期計畫，另有「提升大學基礎教育計畫」其中通過初審的計畫，便有幾個是關於人文電腦輔助教學計畫，如台北大學的「提昇鄉土資源基礎課程設計：三鶯地區史料、古蹟、文物、藝術之數位化輔助教材設計」、文化大學「新發現中國古代文獻課程開發」可見人文學門的教學，已逐漸引用文獻數位化的成果以及資訊網路帶來的便利性。

## 肆、未來的展望

台灣文獻數位化發展迄今已經有十七年，數位化文獻的數量日益累積，數位化的文獻形式日益多元，參與的人員日益增加，資料庫的功能日益提升擴充，參與的單位從個別單位自發性的進行到國家單位有計畫的規劃執行，資料的格式從各自的摸索到訂定統一的標準，相關人才的養成從各自培養研究助理到計畫內有計畫的培養管理人才、大學院校成立相關科系培育人才，至此為止，台灣文獻數位化邁入一個嶄新的時代。

文獻數位化的成果建置於國際學術網路上，所面對的使用群便是全球的使用者，典藏於台灣的中國典籍，只是所有中國文化資產的一部份，文獻數位化的工作，除了台灣之外，中國大陸，乃至於全球收藏中國文獻的圖書館、博物館，都在進行同樣的工作，因此未來台灣文獻數位化的發展，除了持續在數量上增加，種類上擴充、技術上提升、應用上落實，研究與教學上並重，相關人才培育上紮根，還需注意到數位化典藏標準規格的建立以及後設資料格式的建立，以便於與全球各地中國古籍文獻資料數位化的成果接軌。

在數位化的內容方面，建立典藏文獻的後設資料之後，宜進一步建立文獻內容的標誌，並與XML技術結合，以XTM（主題圖，XML Topic Maps）建立文獻內容的知識結構，期使文獻數位化的成果能藉助資訊科技的優勢，展現不同的知識結構方式，拓展新的研究方向與研究領域，使中國古籍在資訊科技的協助下以新的風貌重現光華。

附註：

- 一、引自拙作〈台灣地區中國古籍文獻資料數位化的過程與未來的發展方向〉，收錄於《五十年來台灣人文學術研究叢書---文獻學與圖書資訊學》，台灣學生書局四十週年紀念叢書---圖書文獻類，89年12月出版。
- 二、引自謝清俊著，〈從全文資料庫到數位典藏---中央研究院的發展經驗談〉，香港中文大學新亞書院五十週年紀念---21世紀中文圖書館學術會議，1999年9月30日。
- 三、引自謝清俊編撰，《國家典藏數位化計畫規劃案結案報告》，行政院科技顧問組委辦，中央研究院資訊科學研究所執行，2000年12月20日出版。
- 四、引自自由時報黃國禎報導，〈地方到中央，一網打盡---全國文化資料庫，文建會推動數位化〉，2001年4月5日自由時報40版藝術特區。



- 五、引自民生報于國華報導，〈規畫財產權保障並研擬回饋機制〉 2001年4月12日，民生報 47版文化新聞。
- 六、引自陳昭珍主編，《中文詮釋資料（Metadata）格式彙編》，國家圖書館，2000年12月初版，莊芳榮序言。
- 七、引自國立台北大學中國語文學系網站之本系簡介，網址：  
<http://www.nchulc.edu.tw/dcll/submenu1.htm>
- 八、引自國科會數位博物館計畫網站之計畫簡介。網址：  
<http://dlib1.sinica.edu.tw/OpenQuery/projectintro/index.html> 九、
- 九、僑務委員會所建立的華文教育中心，網址：<http://edu.ocac.gov.tw/index.htm>
- 十、下一代資訊通訊網路尖端技術及應用--網路教育園區及其社會影響研究，設立「卓越計畫網路教育園區」網站，網址：<http://mozilla.hss.nthu.edu.tw/teens/>