

行政院國家科學委員會專題研究計畫成果報告

電子圖書館文獻與博物館藏品數位化技術與應用研究-總計畫  
A study on the technology and applications of the digital library and museum

計畫編號：NSC88-2213-E-002-028

執行期限：87/08/01~88/07/31

主持人：歐陽彥正

處理方式：

可立即對外提供參考

，一年後可對外提供參考

，兩年後可對外提供參考

(必要時，本會得延展發表時限)

執行單位：國立台灣大學 資訊工程研究所

中華民國八十八年十一月

電子圖書館文獻與博物館藏品數位化技術與應用之研究

計畫編號：NSC-88-2213-E-002-028

計畫期限：87/08/01~88/07/31

主持人：歐陽彥正 台灣大學資訊工程學系

一、中文摘要

本計畫的研究目標涵蓋建立大型電子圖書館與博物館各個層面的重要議題。成功的建立一個電子圖書館與博物館的關鍵，除了資訊計畫外，尚需有豐富的內容，並將這些內容，根據使用者之資訊需求予以組織，以提供使用者一個極為人性化的搜尋與檢索界面。因此本計畫的成員包含來自資訊工程學，圖書館學及歷史學的研究人員，組織一個完整的團隊。分別就建立大型電子圖書館與博物館所涉及的

- ˘ 資訊系統設計
- ˘ 使用者行為分析與便易使用環境之建立
- ˘ 資料的整理與組織

三個層面的問題深入研究。下圖所示為本研究群的組織架構：

資訊系統設計				
子計畫一	電子圖書館與博物館資料管理及儲存系統設計之研究	歐陽彥正	台大資訊工程學系暨研究所	教授
子計畫二	台灣史料中知識擷取與覽讀環境之研究	項潔	台大資訊工程學系暨研究所	教授
子計畫三	分散式電子圖書館與博物館系統之整合檢索研究	陳昭珍	輔大圖書資訊學系	副教授
子計畫四	電子圖書館與博物館中影像處理	傅楸善	台大資訊工程學系暨研究所	副教授

使用者行為分析與便易使用環境之建立				
子計畫五	台灣史料使用者資訊需求與搜尋行為之研究	林珊如	台大圖書館學系暨研究所	副教授

資料之整理與組織				
子計畫六	台灣史料整理與數位化	吳密察	台大歷史系暨研究所	副教授
子計畫七	台灣史料數位化資源組織之研究	陳雪華	台大圖書館學系暨研究所	教授

英文摘要

The main purpose of this research project is to investigate the key issues in the design of digital libraries and museums. In order to create an outstanding digital library and museum, it is essential to collect a large quantity of distinguished contents. In addition, the collected contents must be carefully organized to facilitate information retrieval. Accordingly, the research team of this project includes faculty members from the following departments: computer science, library science, and history. The teamwork has been quite effective. The research team completed the first version of National Taiwan University Digital Library and Museum, (NTUDLM) in October 1997 (URL : <http://ntudlm.ntu.edu.tw>). The experiences that we accumulated from developing the NTUDLM help us identify the key issues in designing a digital library

and museum. In this project, we conducted comprehensive studies on

- ˆ design of the information system.
- ˆ Analysis of users behavior and development of friendly user interfaces.
- ˆ Analysis and organization of digitized contents.

## 二、計畫緣由與目的

觀察電腦科技約五十年的發展可以歸納出下列幾個階段。在第一階段，以大型主機為核心的電腦科技主要被應用在科學、工程、軍事與特定商業的計算上。這個時期，只有電腦專業人士才具備操作與使用電腦的技能。到了第二階段，個人電腦與工作站興起後，一般人士才逐步借用電腦完成部份工作。但在這個階段，電腦對一般人而言仍只是輔助性的工具，而非生活或工作的重心。時至今日，由於網際網路的興起，創造了新一代的通訊方式。電腦科技提供的不再僅限於計算與資料的存取。進一步的，電腦已成為人們獲取資料的重要管道，而人類社會也將進入全面資訊化。

由於電腦即將成為文明社會中獲取資訊的重要管道，『資訊提供站』的建構便成為這整個發展步驟中的關鍵工作。這也就是為何美國國科會 (National Science Foundation) 於近來推動『電子圖書館』(或『數位圖書館』initiative)的主要原因。在參考過國外目前進行中的電子圖書館研究計畫及思考網際網路這個新媒體的特性之後，我們歸納出三個層面的研究重點。這三個層面是

- ˆ 資訊系統設計
- ˆ 使用者行為分析與便易使用環境之建立

### ˆ 資料的整理與組織

以下就這三個層面的問題加以說明。並由此引導出本群體計畫的分工合作的架構。

#### (1) 資訊系統設計-

電子圖書館與博物館資訊系統包括三個主要的子系統：資訊管理及儲存子系統，檢索服務子系統，及數位化輸入及編輯子系統。針對這三個子系統，在本計畫中規劃了四個子計畫，研究其中關鍵議題。子計畫一“電子圖書館與博物館資料管理及儲存系統設計之研究”顧名思義是探討資訊管理及儲存子系統設計的課題。檢索服務子系統部份則有兩個子計畫。子計畫二“台灣史料中知識擷取與覽讀環境之研究”，是探討功能強大的檢索系統的設計。尤其是如何自動擷取藏於電子圖書館與博物館資料中的知識，特別是一些靠人力很難或幾乎不可能分析得到的知識。子計畫三“分散式電子圖書館與博物館系統之整合檢索研究”主要研究如何提供使用者一個可在網路上同時檢索數個電子圖書館內容的使用環境。至於第三個子系統“數位化輸入及編輯子系統”，主要是一些 authoring tools。但是在本計畫中由於內容藏品多為古文書及器物，因此引伸出一項重要的課題，即如何運用數位影像處理的技術，強化這些年代久遠文書及器物的視覺效果。這是子計畫四“電子圖書館與博物館中影像處理”的研究主題。

#### (2) 使用者行為分析與便易使用環境之建立-

使用環境之便易性是當今任何資訊系統設計的主要議題之一。欲建立一極便易之使用環境當然必須由研究使用者

需求及行為開始。這便是子計畫五“台灣史料使用者資料需求與搜尋行為之研究”的研究主題。

### (3) 資料的整理與組織-

本計畫數位化的材料是以台灣大學所收藏的古文書及器物為主。而欲在電子圖書館與博物館中針對這些文物進行檢索，便必須先做“feature/attribute extraction”的工作。由於這些文物都具備人文與歷史背景，因此 feature/attribute extraction 的工作必須由具專業知識的研究者來做。例如一件原住民所使用的祭典器物，其上的圖案可能隱藏著其族人相傳的神話。這種解讀工作，不可能由電腦做自動辨認，而必須由專家來作。這部份的工作是子計畫六“台灣史料整理與數位化”的研究主題。

至於專家所解讀出來的資訊，必須配合使用者的資訊需求與檢索行為模式，已一個有組織的方式記錄於電子圖書館與博物館中。這部份的工作是子計畫七“台灣史料數位化資源組織之研究”的研究主題。

## 三、研究方法與成果

以下就本群體計畫各子計畫之成果逐一說明：  
子計畫一“電子圖書館與博物館資料管理及儲存系統設計之研究”

本子計畫的第一年針對 text, image, graphics 等形態資料為主體的素材，完成電子圖書館/博物館中 repository system 及多階層儲存系統的設計。具體成果條列如下：

- Repository system API

- Indexing structure 設計
- Contents manager 設計
- 多階層式儲存系統設計

## 子計畫二“台灣史料中知識擷取與覽讀環境之研究”

本子計畫之成果如下：

- } 建立 meta-metadata 的 methodology 與研究針對子計畫九所發展的 metadata 建立有需求的 meta-metadata。
- } 建立適合於在古文書中 metadata 上做知識擷取的 data mining 方法。
- } 發展 dynamic encyclopedia 之理論，與開發所需輔助工具。

## 子計畫三“分散式電子圖書館與博物館系統之整合檢索研究”

本計畫完成之工作項目如下：

1. 了解目前已開發出之整合性資訊檢索模式之種類
2. 了解各種整合性檢索模式的系統架構、功能、優缺點、長遠性、延展性。
3. 了解 Java、XML、SGML、Metadata 與整合性檢索的關係。
4. 了解 CIMI Profile 及 Z39.50 Profile for Access to Digital Collections 以及 CIMIzit Agreement 相關規定。
5. 了解台灣史電子圖書館使用 Z39.50 標準做整合性檢索的可行性。
6. 了解曾設計過整合性檢索程式或對整合性檢索有深入之專家學者之設計心得及經驗。

## 子計畫四“電子圖書館與博物館中影像處理”

本子計畫完成的工作項目如下：

(一) 在影像提昇技術方面完成：

1. 區域長條圖指定法
2. 空間低通濾波法
3. 邊緣偵測法

(二) 在影像復原技術方面完成：

1. 幾何轉換，如：亮度插入。
2. 反濾波方法

#### 子計畫五“台灣史料使用者資訊需求與搜尋行為之研究”

本子計畫達成下列目標：

1. 深入地了解平埔研究人員及一般民眾的資訊搜尋與使用行為。
2. 提供發展本土化資訊行為理論之知識與實證資料。
3. 辨識台灣研究之重要資源，並建立台灣歷史與文化文獻資料的特徵，以支援台灣史料整理與數位化。
4. 提出國內相關本土化資料之數位化資訊組織，應有之欄位與有效之檢索點。
5. 引導出支援資訊搜尋行為的有效資訊檢索策略與技術，輔助索引點與索引辭彙之研究。
6. 提出數位化文獻資源設計原則，如覽讀環境之建構，供國內相關電子圖書館/博物館系統作為界面設計之參考。

#### 子計畫六“台灣史料整理與數位化”

本子計畫成果如下：

- (一) 以台灣大學圖書館藏《淡新檔案》《岸裡大社文書》《伊能文庫》等寶貴台灣史資料

為基礎，建構一個質、量均佳的電子圖書館/博物館，提供全民使用。

(二) 研發出著錄各種類型之歷史資料的 metadata 欄位，使各種類型史料之間可以透過機連絡。

(三) 在建立資料庫的過程中，累積性地編成台灣史相關的辭典，例如：類音辭典、同意辭典。

(四) 建立中文世界之電子圖書館/博物館標準。

(五) 以此計畫之經驗，提供台灣研究資源全國性整合之參考。

#### 子計畫七“台灣史料數位化資源組織之研究”

本研究計畫完成之工作項目：

1. 探討國內外電子圖書館資訊組織之法，分析其利弊得失。
2. 探討幾種以發展出來的 metadata，分析其利弊得失及未來可能之發展以供國內參考。
3. 調查分析台大及其他單位現有珍藏之文獻(物)特性以及使用者之資訊檢索需求，建議描述這些電子文獻之 metadata 與資訊組織模式。
4. 訪問學者專家及使用者，實地了解其資訊需求及使用情境。
5. 擬定文書類 metadata 並設計工作單，提供相關單位實際輸入資料，測試後並加以修正。

#### 四、結論

本計畫的綜合效益，可以分兩方

面來說，一方面是建立國內在建構電子圖書館與博物館的科技能力；一方面是在保存文化資產與提升台灣史之研究、教育環境。

我們相信經過本計畫兩年的努力與經驗累積，我們將可完全掌握電子圖書館/博物館系統設計上的相關技術。由於電子圖書館與博物館的建構式資訊科技領域新興的研究主題，因此本計畫的執行，對建立國內資訊學界這方面的研究實力將有相當大的助益，亦將產生許多國際上相當領先的研究成果。這些成果不僅有學術上的貢獻，同時也具有產業的價值。我們可以預期在未來資訊化的社會當中，透過網際網路做資料搜尋無疑次提升生活品質與工作效率不可或缺的一環。而電子圖書館與博物館正是這種社會活動中最重要資訊來源。因此建構電子圖書館與博物館所涉及的資訊科技將是未來資訊產業重要的關鍵技術。此外，基於我國在全球華人社會中的地位與科技實力，實應積極朝向成為華人文化科技中心的目標邁進。本計畫的執行，將有助於建立我國在中文電子圖書館與博物館的研究地位。使得這項成為華人文化科技中心的目標更能及早實現。

本計畫所將存入電子圖書館與博物館的資料為台灣史相關之珍貴史料、文物。這些本土的史料、文物，是從事台灣史研究的珍貴基礎資料。長期以來，台灣史研究專家以傳統的方式，耗盡畢生精力，亦只能掌握、利用其中的一部份，如今借助科技之助，可以全面性地、快速獲得被某些機關秘藏之資料，而且由工程與資料管理的專家提供進一步使用這些資料的可能性。這乃是資訊科技與台灣史學術研究配合，建立新時代優異研究環境的範例。藉助數位化處理與網際網路的傳

播，也可以使這個計畫的成果提供為教育、學習之資源，而且將使台灣史研究具有國際化的基礎，使台灣史研究這個本土研究主題，可以參加國際學術討論。

## 五、參考文獻

- } Jian-Hua Yeh and Yen-Jen Oyang, "Content and Knowledge Management in a Digital Library" Accepted by the Journal of American Society for Information Science, Special Issue on Digital Library, (Publisher: American Society for Information).
- } Jian-Huan Yen, Jia-Yang Chang, and Yen-Jen Oyang, "Exploiting of Information Processing Technologies to Extend Digital Library Applications", Journal of Library and Information Science, Vol.25, No.1, pp.1-11, April 1999.
- } Chung-Chen Lai, Juan-Hua Yeh, and Yen-Jen Oyang, " Toward Automatic Generation of Geographic Thesaurus through Data Mining in a Geographic Information System",
- } Chien-Kang Huang, Nai-Wen Lin, and Yen-Jen Oyang, " A Search Engine for Multiple-Version Chinese Buddhist Scriptures", Proceedings of the Second Asian Digital Library Conference, Taipei, Taiwan, November 1999.
- } Jian-Hua Yeh, Hsiu-Ping Lin, Chun-Yuan Shao, Yi-Chin Tu, and Yen-Jen Oyang, "System Design of the National Taiwan University Digital Library and Museum", Proceedings of the First

- Asia Digital Library Workshop, Hong Kong, August 1998.
- } 陳雪華。「博物館典藏品數位化資訊組織之探討」。大學圖書館 3 卷 1 期 (出版中)。
  - } 陳昭珍。「中文字料資訊 Z39.50 Profile 的設計」。大學圖書館 3 卷 1 期 (出版中)。
  - } 陳雪華。「Metadata 在電子圖書館所扮演的角色」。國立成功大學圖書館館刊 3 期 (民國 88 年 4 月), 頁 16-24。
  - } 陳光華。「資訊檢索技術之核心」。大學圖書館 3 卷 1 期 (出版中)。
  - } 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。
  - } 陳昭珍。「控制詞彙與非控制詞彙主題索引問題之探討」。圖書館學刊 13 期 (民國 87 年 12 月), 頁 103-128。
  - } 陳昭珍。「從使用者需求與文獻特性看圖書館資訊組織模式發展趨勢」。大學圖書館 3 卷 1 期 (出版中)。
  - } 陳嵩榮、陳昭珍。「XML/RDF Metadata 模式之發展與實作」。華文書目資料庫研討會論文集。(台北: 國家圖書館, 民國 88 年 8/30-9/1)
  - } Chang, Shan-Ju Lin, Mu-Hsuan Huang. "Digital Libraries: User Studies." In: Ching-chin Chen, ed. IT and Global Digital Library Development (Proceedings of NIT 1999: The 11th International Conference on New Information Technology, August 18-20, 1999, Taipei, Taiwan.) West Newton, Mass.: MicroUse Information, 1999, pp. 33-42.
  - } Chen, Hsueh-hua, Chen, Chao-chen and Chen, Kuang-hua. "Metadata Interchange for Chinese Information." In: Ching-chin Chen, ed. IT and Global Digital Library Development (Proceedings of NIT 1999: The 11th International Conference on New Information Technology, August 18-20, 1999, Taipei, Taiwan.). West Newton, Mass.: MicroUse Information, 1999, pp. 65-74.
  - } Chen, Hsueh-hua. "METADATA: The Foundations of Resource Description for Digital Libraries/Museums" "Proceedings of 1999 EBTI, ECAI, SEER & PNC Joint Meeting, January 18-21, 1999, Taipei, Taiwan, pp. 83-107.
  - } Chen, Kuang-hua and Wu, Chien-tin. "Automatically Controlled-Vocabulary Indexing for Text Retrieval," Proceedings of the 12th Research on Computational Linguistics Conference, (ROCLING99), Hsinchu, Taiwan, ROC, 1999, pp. 171-185.
  - } 陳昭珍。「資訊檢索服務與協定」。

NII 資訊及通信標準應用研討會論文集。(高雄；台北：經濟部標準檢驗局主辦；資訊工業策進會協辦，民國 88 年 4 月 29-30 日(高雄)；5 月 18-19 日(台北市)

- } 陳昭珍。「政府出版品電子化格式與作業標準」。電子出版品電子化研討會論文集。(台北市：行政院研究發展考核委員會,民國 88 年 3 月 3 日)
- } 陳昭珍。「控制詞彙、非控制詞彙與主題索引的未來」。2000 讀者服務研討會。(台北市：實踐大學圖書館，民國 88 年 1 月 8-9 日)