

數位圖書館/博物館評鑑指標建構探討

吳美美* 林珊如** 黃慕萱*** 葉乃靜****

摘 要

進入二十一世紀之前，全世界進步中國家競以發展太空科技、生物科技、建立美輪美奐的國家圖書館、發展虛擬化的數位圖書館/博物館來彰顯國力、彰顯跨足新世紀的能力。在前述各項建設之先，必須先有詳細的藍圖規畫、模擬、測試，以及預估成效。尤其系統開發的不同階段，需要進行適當的評估，例如，系統設計之前的「需求評估」、系統雛形設計完成的「形成性評估」、以及系統開發完成的「總結評估」，可以對系統開發的投資報酬作最有效的管理。本文說明數位圖書館評鑑研究的重要、列舉世界重要數位圖書館計畫的評鑑研究、探討數位圖書館評鑑研究的相關議題，提出一個系統評鑑的理論模式，最後是結論。

關鍵字：數位圖書館/博物館評鑑 檢索系統評鑑 評鑑研究 使用者評鑑 使用者研究 評鑑指標

*國立臺灣師範大學社會教育學系副教授 (Graduate Institute of Adult and Continuing Education, NTNU, Associate Professor, meiwu@cc.ntnu.edu.tw)

**國立臺灣大學圖書資訊學系副教授(Department of Library and Information Science, Associate Professor, NTU, sjlin@ccms.ntu.edu.tw)

***國立臺灣大學圖書資訊學系教授(Department of Library and Information Science, Professor, NTU, mhhuang@ccms.ntu.edu.tw)

****世新大學圖書資訊學系講師 (Department of Library and Information Studies, Shih Hsin University, Instructor, ncyeh@cc.shu.edu.tw)

Understanding Digital Library/Museum Evaluation

Mei-Mei Wu*, Shan-Ju Lin Chang**, Mu-Hsuan Huang***, Nie-Ching Yeh****

Abstract

Digital Museum Program launched by National Science Council (NSC) has completed its first phase auditing at the end of 1999. The evaluation programs for DL/M at the different levels for the various purposes are highly recommended. The typical spots for DL/M evaluation include system analysis phase, prototype stage (formative evaluation), performance evaluation (summative evaluation), and for the resources management purposes -- cost-effectiveness evaluation. This paper discusses the important functions of evaluation at the different stages of system design, reviews the major DL/M evaluation projects, delineates the critical issues of DL/M evaluation program. And finally, a theoretical framework for DL/M evaluation program is suggested.

Keywords: DL/M evaluation, user study, evaluation criteria

壹、前言 數位圖書館/博物館計畫為什麼需要評鑑研究？

「評鑑研究」是發展數位圖書館/博物館過程中不可省略的重要部分。在數位圖書館/博物館發展相關研究報告中，可以發現世界上重要的數位圖書館/博物館的開發案例，都將評鑑研究、評鑑報告乃至評鑑綱領視為計畫的一部分。究其原因，實乃評鑑研究結果所反應的系統效益或使用者的看法，是建置完善數位圖書館/博物館的主要參考資料。尤其數位圖書館/博物館還在開發的起始階段，仍需面對許多理論和技術上的挑戰，評鑑研究有助於釐清這些理論和技術上的困難。

許多數位圖書館/博物館的研究者(如 Belkin, 1984 ; Kilker & Gay,1998 ; Pejtersen, Dunlop & Fidel, 1999 等)都指出數位圖書館/博物館的開發成敗，除了應注意儲存、顯示、轉換、列印等技術層面的問題之外，使用者需求、內容層次的設計都是不能忽略的重要課題。數位圖書館/博物館為了達到資訊無遠弗屆傳播的理想，需考慮廣泛的虛擬使用者，使用者的資訊行為是什麼？其資訊需求是什麼？進一步而言，數位圖書館/博物館在規畫建置時，首先便要進行使用者需求分析，使用者需求分析及使用者導向評鑑，是數位圖書館/博物館發展成功的重要關鍵。而從數位圖書館/博物館贊助單位來說，系統開發的計畫更須評鑑，其評鑑的目的，是管理的層次，超越系統設計者僅為求改進系統功能的單純目的，也就是更關心整體系統的投資報酬和效益。這些效益的考量，除了使用者個人的層次之外，還有社群的集體效益和社會共同效益，或甚而對改善整體人類生活的層次。

世界著名的數位圖書館先導研究計劃都重視評鑑研究。Bishop (1995) 參與美國「數位圖書館先導計畫 I」(DL-Initiative I)的評鑑研究工作，曾將「數位圖書館先導計畫 I」評鑑小組的工作目標訂定為針對發展中的檢索機制、收費計畫和其他數位圖書館系統特色及功能，收集使用者回饋意見、分析實驗階段使用情形(tested use)、使用者滿意度和影響滿意程度的原因、了解使用與不使用的理由，以對系統實驗階段之成敗有更完整的了解、了解發展中的資訊基礎建設，並對圖書館工作、資訊傳播和學習有更深入的認識、發展以使用者為中心的數位圖書館評鑑研究方法等。這些評鑑的目的勾勒數位圖書館在發展階段，除了相關技術的開發之外，在使用層面所應重視的相關課題。

英國「數位圖書館研究計畫」(ELIB Project) 出版一份完整的評鑑手冊「數位圖書館評鑑程序」(Guidelines for ELIB Project Evaluation) (Kelleher, Sommerlad & Stern, 1996)。該手冊建議每一個個別的數位圖書館/博物館計畫都需要進行評鑑研究，共分為 11 章，包括簡介、手冊的內容說明、系統介紹(說明評鑑系統之特色，定義數位圖書館，以便有助於評鑑研究參數的擬訂，並有助於讀者閱讀接下來的評鑑大綱)、評鑑設計的一般原則(包括評鑑問題和評鑑者的目的、與其他資料蒐

集活動的整合、使用者在評鑑研究中的角色，此外評鑑設計原則若與評鑑目的不同，予以說明）實施評鑑需知、善用評鑑結果、評鑑計畫的研擬（說明評鑑研究採用的方法、評鑑的項目、以及評鑑的指標）評鑑技術支援、年度報告（說明評鑑報告與年度報告的關係）其他數位圖書館評鑑活動（與該評鑑計畫相關的其他數位圖書館計畫應一并描述）最後是參考文獻（詳列對評鑑研究有助益的相關參考文獻）。英國的「數位圖書館研究計畫」(ELIB Project)視評鑑為數位圖書館系統開發的一個部分。

評鑑研究是數位圖書館/博物館計畫的一部分，評鑑研究的目的包括了解及改善系統功能、了解及改善使用者檢索效益、了解及善系統分析級系統設計的有效進行，在管理的層次，利用系統化的方法提升數位圖書館計畫的品質，促進其效益。本文以下分別討論國內外數位圖書館/博物館評鑑重要實例、評鑑指標、評鑑的重要議題、最後提出結論和建議。

貳、國內外重要數位圖書館/博物館評鑑實例討論

國內外有幾個重要的數位圖書館/博物館評鑑實例，本文選擇介紹「美國記憶」(American Memory, AM)、「美國數位圖書館計畫系統」(The Making of American Digital Library Project, MoA)、美國政府資訊指引服務系統(The Government Information Locator Service, GILS)、加州大學柏克萊分校的環境規畫與地理資訊系統(University of California at Berkeley: Environmental Planning and Geographic Information Systems)以及台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統等五例。

一、「美國記憶」(American Memory, AM)

美國國會圖書館的「美國記憶」(American Memory, AM)計畫，是美國國會圖書館數位圖書館先導計畫，1990年開始發展，主要的目的是將美國國會圖書館的館藏資源，以多媒體電子形式呈現，並向全球傳播。其內容強調具有美國歷史和文化的研究價值，尤其是稀少而珍貴的資料，特別予以收集和收藏。為求廣泛的傳播，每件實體資料都以機讀編目格式(MARC)詳細編目，並利用關鍵字和簡單的布林邏輯檢索(Polly & Lyon, 1995, p.51)。

美國國會圖書館為了解該系統的主要使用對象，在1989年和1990年，該計畫已分別進行兩次小型的評鑑研究，在1991年4月成立「國會圖書館使用者評鑑團隊專案計畫」(The Library of Congress Special Project User Evaluation Team)，在1991至1993年間進行大規模的使用者評鑑調查。為能夠反應該系統所有可能的使用者，美國國會圖書館慎選評鑑對象，其評鑑的圖書館範圍包括各級學校(K-12)、

大學和學院、公共圖書館、州立圖書館、專門圖書館等共四十餘個圖書館。該計畫的主要評鑑目標有五項：「美國記憶」的使用者是誰？那些館藏被使用？資料如何被使用？那些支援資料（如展示的實用性、館藏資訊、線上輔助、紙本資料）被使用？系統可使用性(Useability)如何？(The Library of Congress, 1993, chapter1, p.1 of 3)

自 1991 年秋季起，總共有四十四個圖書館單位自願參與該研究。該研究利用問卷、電話訪談、當面訪談、觀察使用者瀏覽網站、以及查詢記錄分析等方法蒐集了量化和質化的資料。量化資料包括 1801 份問卷、44 個機構中收到 21 個機構的查詢記錄；質化資料則有 55 位職員及 121 位使用者的個別訪談、電話訪談、資料庫瀏覽(觀察)等，並於 1993 年 6 月完成研究分析(The Library of Congress, 1993, chapter1, p. 2 of 3)。

該評鑑研究有幾個重要的發現：「美國記憶」主要的使用者是中等學校，但隨著內容成長，該計畫希望其使用者層面能擴展，甚至能支援高等教育群之學術研究的進行；最常被使用的資料是照片館藏 (photograph collections)；而教育人員如何使用主要資源，應再深入了解研究。此外也發現該系統的主要支援系統（如線上求助功能）不是十分令人滿意，應包含完整的系統使用說明；在評鑑研究中發現使用者最關心的是資料庫內容，系統應加強資料庫收錄範圍和內容描述，以利使用者掌握文獻的出處及完整性；考慮使用光碟（CD-ROM）系統的可能性；簡化介面並加強蒐尋引擎的功能；整合各類型資料，以滿足不同使用者的需求（The Library of Congress, 1993, chapter 6 ）。

二、美國數位圖書館計畫系統（The Making of American Digital Library Project, MoA）

「美國數位圖書館計畫系統」(The Making of American Digital Library Project, MoA)由康乃爾大學（Cornell University）和密西根大學（University of Michigan）兩校共同發展，美國「安德魯·美隆基金會」（Andrew W. Mellon Foundation）經費贊助。該計畫初期選擇兩校蒐藏的十九世紀美國期刊和論文，將一百五十萬頁的資料數位化，再透過全球資訊網傳遞給使用者，提供數位資料檢索（Kilker & Gay, 1998, p.62）。

該計畫實施評鑑研究，以了解目標達成的情況。該項評鑑研究應用「使用者接受度」（user acceptance）和「系統使用性」（usability）兩項評鑑指標，用以檢驗該系統的效益（Kilker & Gay, 1998, p.62）該研究為進一步了解數位資料影響研究和教學的情況，採用訪談、開放式問卷、焦點團體等進行資料收集，對象包括圖書館員、大學生和研究生。研究發現，使用者認為理想的數位圖書館應該有下列

的功能：具備有效力的檢索工具、快速檢索、真實情境瀏覽（realistic viewing）、檢索方式較應紙本翻頁或使用微縮資料容易、瀏覽及航行的功能具有微縮資料的特色等（Kilker & Gay, 1998, p.63）。

根據 Kilker & Gay (1998, p.64) 該評鑑研究也發現不同的使用社群，如圖書館員、系統設計者、老師和學生等，對系統功能有不同的要求，譬如系統設計者著重個人航行習慣、個人化參考工具、介面顯示特色、容易操作、排序、修正工具等層面，圖書館員著重系統一般特色、檢索功能、版權、有效性和適應性、列印和下載、載入等功能。Kilker & Gay (1998)曾評比系統設計者和圖書館員對評鑑項目的不同優先順序(如表一)。系統評鑑時，應注意不同使用者(群)對系統效能有不同的觀點。

表一 系統設計者和圖書館員對評鑑項目的優先順序(Kilker & Gay, 1998, pp.60-70)

優先順序	設計者	圖書館員
1	航行習慣	一般特色
2	個人化參考工具	檢索
3	文獻顯示習慣	文獻顯示習慣
4	介面顯示特色	版權
5	容易操作	航行
6	檢索和瀏覽工具	有效性和適應性
7	排序	列印和下載
8	修正工具	檢索結果顯示
9	研究檢索工具	載入
10	人的面向	
11	檢索選擇	
12	視覺/人性觀點	
13	分享	

三、美國政府資訊指引服務系統(The Government Information Locator Service, GILS)

美國政府資訊指引服務系統(The Government Information Locator Service, GILS)，發展於 1994 年 12 月，主要依據美國聯邦政府公布的行政命令 95-01 (The Office and Management Bulletin (OMB) 95-01)，該系統的主要目的在協助使用者於網路環境下，能順利找到政府的資訊資源（Moen & McClure, 1997）。

1995 年 12 月，John Carlin 在「美國政府資訊指引服務委員會」建議進行該系統評鑑研究，以了解該系統符合使用者資訊需求的情形，做為改善系統之參考。評鑑研究於 1996 年 9 月至 1997 年 3 月進行，具體的研究目的包括：了解美國政

府資訊指引服務(GILS)的服務對象、對大眾檢索政府資訊的影響、對民眾了解政府機構進展情形是否有所助益、以及該系統如何扮演資訊資源的管理工具等。

該研究採用的評鑑方法包括資料庫瀏覽(site visits) 焦點團體(focus groups) 調查法(survey) 美國政府資訊指引服務(GILS)記錄內容分析(GILS record content analysis) 線上使用者評鑑 (online user assessments) 查詢記錄分析 (web server transaction log analysis) 政策及文獻探討 (policy and literature review) (Moen & McClure, 1997)。

研究結果包括目的和使用、改善美國政府資訊指引服務(GILS)、美國政府資訊指引服務(GILS)與其他資訊處理功能的關係、加強使用者對美國政府資訊指引服務(GILS)的認知等四方面,以下進一步說明(一)目的和使用(purpose and utility) 方面,使用者對系統的目的、任務和使用方式並不了解;系統的建置缺乏行政部門的合作;各機構的文化和任務不同,對該項服務的反應和期望也不同;美國政府資訊指引服務(GILS)的效益、責任並不明確。(二)改善美國政府資訊指引服務(GILS)方面,網路化資訊檢索環境已提升美國政府資訊指引服務(GILS)的重要角色;美國政府資訊指引服務(GILS)以機關為中心,非以政府為中心;美國政府資訊指引服務(GILS)的詮釋資料(metadata) 訂定並不容易;美國政府資訊指引服務(GILS)記錄的更新和維護非常有限;美國政府資訊指引服務(GILS)記錄的資料項目,並未獲得一致的認同;美國政府資訊指引服務(GILS)記錄中包含不同類型的資源;使用者對美國政府資訊指引服務(GILS)的回應並不積極;美國政府資訊指引服務(GILS)記錄的展現方式並不相同;使用者指引並不妥當。(三)美國政府資訊指引服務(GILS)與其他資訊處理功能的關係方面,美國政府資訊指引服務(GILS)與政府機關資源間的關係不明確;美國政府資訊指引服務(GILS)與資訊自由法案(Freedom of Information Act, 簡稱 FOIA) 的關係不明確。(四)加強使用者對美國政府資訊指引服務(GILS)的認知方面,缺乏美國政府資訊指引服務(GILS)推廣和教育計畫;美國政府資訊指引服務(GILS)的潛在使用者對該系統並不熟悉;美國政府資訊指引服務(GILS)的使用情形很有限。這些評鑑研究結果顯示該系統仍有很大的改善空間,對使用者和系統效益而言是值得鼓勵和讚賞的研究(Moen & McClure, 1997)。

四、加州大學柏克萊分校的環境規畫與地理資訊系統(University of California at Berkeley : Environmental Planning and Georgraphic Information Systems)

加州大學柏克萊分校的「環境規畫與地理資訊系統」(University of California at Berkeley : Environmental Planning and Georgraphic Information Systems) 是「美國數位圖書館先導計畫」(Digital Library Initiative) 六大子計畫之一,主要的贊助者是美國 NSF/NASA/ARPA 三個單位(Van House, 1995) 該計畫主要的目標在蒐

集、典藏、組織數位化資訊，並透過網路提供高親和性的檢索功能。

「環境規畫與地理資訊系統」是分散式、以滿足使用者工作上的需求為主的大型數位化資訊系統，為擴大使用率並遵循以使用者為中心的系統設計原則，該系統於發展之初，曾分別就專家和使用者兩方面進行評鑑研究。

使用者需求評鑑研究主要的樣本為「水資源部門」(Department of Water Resources) 的研究人員、科學家、工程師、作家、教育人員等。評鑑研究的對象除了上述人員外，更擴及一般有興趣的社會大眾。研究方法包括深度半結構式訪談、焦點團體訪談、語言分析 (protocol analysis)、使用者行為觀察、文獻分析等方法。在以使用者為主的系統設計方面，該研究建議未來評鑑研究的思考方向是如何在異質性高的使用者群中，找出更具親和性的系統設計原則，以反應專家和大眾的不同需求，換言之，該計畫的開發宗旨認為除反應母機構、滿足專業工作上的需求之外，一般大眾的需求也應加以考慮。

五、台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統

受英美等國的影響，國內各單位對於數位圖書館的建置和推展，也不遺餘力，例如國家圖書館、中央研究院、故宮博物院、各大學圖書館等，已先後開始規畫。不過對於數位圖書館的評鑑研究則尚未普及。張嘉彬 (民 88) 針對台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統進行的評鑑研究，可為參考。

張嘉彬(民 88 年)由使用者的角度評鑑台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統的功能、檢索者對系統的整體觀感、以及系統支援學術研究的能力。研究方法包括文獻分析法、問卷法、查詢過程記錄分析法、不介入觀察法及深度訪談法等。依線上問卷調查結果顯示，該系統使用者以男性居多，以 20 歲以下及 20-29 歲年齡層的使用者佔大多數，使用者中七成以上具大專以上學歷，且近七成使用者是學生。就系統使用目的而言，高中生以下的學生使用者以課堂作業為主；大專生使用者除了課堂作業外，也為了個人興趣及嗜好；研究生使用者則以從事學術研究為主要目的。

就查詢記錄分析來看，有 30% 的使用者進入首頁後即跳出或關閉視窗；進入首頁者以讀取虛擬博物館的次數最多；使用資料以讀取地圖為最；古文書內容雖豐富，但讀取者並不多。就查詢而言，以關鍵字查詢者取多，很多使用者以自然語言方式輸入查詢敘述。

由訪談資料顯示，使用者認為系統資訊量不足，不能支援其學術研究，使用者的建議包括：增加瀏覽項目之說明、改變瀏覽選單方式、瀏覽後系統能維持上

次瀏覽之款項、合併關鍵字查詢及複合查詢等。

該評鑑研究的結果除了依據研究發現發展出評鑑中文古文書之系統功能簡表外，具體的研究結果如：使用者對系統功能並不滿意，應再增加布林邏輯、切截、縮小範圍查詢、索引典、允許輸入某個區間的時代及接受自然語言方法等功能，檢索結果並有獨立展示畫面、取消多重視窗方式、電子檔建立後應行人工判斷、古文書原件應放置於明顯處、局部放大、檢索結果提供年代排序等功能。此外，系統也要加強輔助說明功能，包括：系統介紹、錯誤訊息顯示、檢索歷史清單、專有名詞解釋、系統需求及常問問題集等。這些研究結果，可提供系統設計者的參考。

以上五例皆以了解使用者（包含潛在使用者）使用者行為，以及使用者對系統功能的意見為主。由於多是研發單位的數位圖書館先導計畫，主要的研究目的皆以讀者為導向，這是合理且可以理解的。至於要了解系統的「表現」（performance），亦即要了解數位圖書館的效能，則需有一套系統化的評鑑指標。以下討論系統評鑑指標的內容。

參、有關數位圖書館 / 博物館的評鑑指標

數位圖書館/博物館評鑑的重要性雖然受到重視肯定，但是對於其中評鑑的項目、評鑑指標和評鑑的方法則仍莫衷一是。由於數位圖書館/博物館是新發展的系統，一方面有技術上的考驗，例如需整合各種不同類型的資訊資源，包括印刷式的文獻、多媒體的聲音、圖像、以及數位化實體資源等，另外資訊的利用是分散式的，使用群廣泛、異質等，使評鑑的範圍和評鑑的方法仍有研議的必要。其中評鑑指標的建立更是系統評鑑最重要的議題之一，從 60 年代以降，評鑑指標的問題就不斷的被討論。以下從系統設計者、中介者和使用者的觀點探討資訊檢索系統、搜尋引擎、網站品質、以及數位圖書館/博物館等評鑑指標的形成。

一、資訊檢索系統

Lancaster 等人曾提及，傳統資訊系統的服務評鑑常用三個標準：費用標準、時間標準、品質考量。品質的考量包括：資料庫含蓋的範圍、回現率（檢索結果的完整性）、精確率（檢索結果的相關性）、新穎性、資料的完整性和正確性。具體言之，評鑑系統要素應包括：可獲取性（acceptability）、容易使用、檢索結果的品質、組織的影響（organizational impact）、可使用性（usability）、含蓋範圍等（Lancaster et al. 1996, p. 59-60）。

Su(1992, p.505) 為找出資訊檢索表現的評鑑指標，由四十個使用者（real users）

和六個專業中介者，檢索四十個問題，利用 DIALOG, BRS, STN 等檢索系統，透過檢索前後問卷、檢索互動過程的錄影，由使用者對檢索結果進行相關判斷，結束後並進行訪談錄音，並以檢索結果的成功率，結合研究結果發現，與檢索成功相關的判斷有下列幾項：使用者對結果完整性的滿意度、整體而言的檢索價值、使用者對結果完整性的信心、檢索面談成功、檢索結果完整的重要性、使用者對精確率的滿意度、檢索者對資料庫使用的知識等，效率和互動、時間也是重要資訊檢索評鑑因素 (Su, 1992, p.509)。

Su (1992, p.505) 同時將影響使用者評鑑的因素一并考慮：如使用者的學位、檢索目的、主題專長、資訊需求的主題領域，並依前此研究的結果，建立了二十個評鑑指標，分屬於五個評鑑標準：

- (一) 相關方面：精確率
- (二) 效率方面：檢索時間、相關評鑑時間、檢索費用
- (三) 效益方面：與所花時間相較，檢索結果價值、與所花精力相較，檢索結果價值、與所花心力相較，檢索結果價值、整體而言的檢索價值
- (四) 使用者滿意度方面：檢索者對需求的了解程度、檢索者對需求目的的了解程度、檢索者對檢索結果完整性的滿意度、檢索者對資料庫使用的知識、對需求能採用合宜的檢索詞、檢索面談、使用者對結果完整性的滿意度、使用者對結果完整性的信心、檢索結果完整的重要性、使用者對精確率的滿意度、精確率對使用者的重要性
- (五) 成功方面：使用者對系統成功的判斷

張嘉彬 (民 88) 列舉 Lancaster、Kiewitt、Courtois、Baer & Stark、Chu & Rosenhal、Dong & Su、黃慕萱、卜小蝶、陳秀美等八個研究的系統評鑑標準，其中「回應時間」為八位研究者共同認同；「精確率」及「檢索結果展示」有七個研究者認同；「資料收錄範圍」、「回現率」、「資料新穎性」有六個研究者認同；「檢索功能」、「成本」有五個研究者認同；「欄位索引情況」有四個研究者認同；「易於使用」、「輔助說明」有三個研究者認同；「使用者研究」、「連結性」及「資料庫建構」有兩位研究者認同；「資料正確性」、「檢索者」、「錯誤率」、「猶豫程度」、「容易連線程度」、「資料庫大小」及「資料庫簡介」等七項都只被一位研究者列為評鑑標準。

張嘉彬 (民 88) 再針對系統檢索功能將 O'Rourke、Piternick、Northover、Cherry、Williamson、Jones-Simmons & Gu、Brinkley & Burke、Baker、康芳菁、李德竹等人的研究，詳列其系統功能之評鑑指標。研究者共同選擇的系統評鑑指標為「檢索功能」、「檢索結果展示」、「輔助說明」等三項；其中有五個研究者都選擇「輸出控制」；有四個研究者選擇「易於使用」及「詞彙協助」等兩項評鑑標準；

有三個研究者選擇「檢索點多寡」；「系統使用資訊」、「回應時間」及「資料收錄範圍」等標準，各有二個研究者選擇；只有一個研究者選擇「容易連線程度」。研究者看法殊異，但評鑑指標的建立殊為不易。

二、搜尋引擎

搜尋引擎評鑑是 1995 - 1996 年間才出現的新領域，因為在索引技術、含蓋資源範圍、檢索結果排序、檢索策略等方面的特色，使其評鑑的方法和評鑑標準不同於傳統的線上資料庫及光碟評鑑 (Dong & Su, 1997, p.69)。Dong & Su (1997, p.68) 指出，評鑑搜尋引擎的方法有三種：蒐集並分析實際上的測試資料集、就簡單的檢索問題予以評鑑、不同搜尋引擎功能比較。他們分析歸納九位學者對搜尋引擎所做的評鑑研究，其中 Lycos 受到最多次評鑑。

Dong & Su (1997, p.70-71) 綜合搜尋引擎的特色，歸納評鑑項目和評鑑內容包括：(一)網頁大小 (size)；(二)資料庫含蓋範圍：是否包括網址 (URL、HTTP)、資訊查詢系統 (Gopher)、檔案傳輸服務 (FTP)、網路新聞討論群 (Usenet Newsgroup)、網頁評論 (Web Page Reviews)、影像視訊 (Image/Video)、圖片 (Graphic) 等；(三)索引欄位方面：包括全文、網址、題名、標題、加權字 (weighted word)；(四)檢索功能方面：包括布林邏輯、關鍵字、片語、主題表、切截、選擇類似網頁、檢索範例；(五)檢索結果排序方面：依相關程度排序；(六)檢索結果顯示方面：包括摘要、大小、標準的呈現方式、詳細的呈現方式、時間反應；(七)更新頻率方面：每天、每週、或不規則等。

Dong & Su (1997, p.77) 進一步提出搜尋引擎評鑑研究中採用的評鑑標準應包括：介面、檢索結果顯示的選擇、精確率、反應時間、資料庫含蓋範圍、串連的有效性、檢索功能、更新頻率、相關排序、索引範圍、資料庫架構、相關程度、結果的完整性、價格、檢索結果的重疊及回現率，而這些評鑑指標和資訊檢索系統評鑑指標互相比較，只有「串連的有效性」是檢索引擎評鑑的獨特項目。

對使用者而言，Dong & Su (1997, p.79) 認為一個好的搜尋引擎應包括下列特質：剖析自然語言，能夠了解使用者的資訊需求、顯示最相關的檢索結果，並予以排序、不能期望使用者會使用布林邏輯或其他更深入的邏輯陳述、不能期望使用者對檢索技能很熟、好的串連效果，指引要達到效果、不要有重複或無效的串連、不要浪費使用者的時間指引至不相關的網站、具親和性的介面、快速的回應時間、回現率和精確率高、提供主題檢索、提供清楚的檢索結果、資料庫含蓋範圍完整且內容常更新。

三、網站品質的評鑑指標

在全球資訊網成為網路資源運用的趨勢後，不論個人或機構都紛紛架設網站，以延伸服務的範圍，因此，網站品質的評鑑也就成了一個愈來愈受重視的主題。對一般大眾而言，也應具備網站品質評鑑的能力，方能找到自己所需的網路資源，不致於浪費時間和人力。Pejtersen, Dunlop & Fidel (1999) 指出以使用者為中心的網路資源評鑑應注意六個問題：網站是否符合使用者特性？網站是否支援策略？網站是否支援工作上所需的認知決策？網站是否支援工作情境？網站是否支援角色扮演及合作性工作？是否在真實環境中實施田野調查？McMurdo(1998, p.193) 也認為評鑑網際網路和全球資訊網，主要的重點在於內容、網站的設計和使用性，使用性主要的焦點則在人機互動上，另有許多研究者針對網站品質提供各種評估指標(如 Druse, 1991 ; Courtois, Baer & Stark,1995 ; Tate & Alexander,1996 ; Discovery Channel School, 1997 ; Goral, 1998 ; 翁慧娟, 1999 等)，擇要說明如後。

Goral (1998) 發展網站評鑑查核表，查核標準及查核項目如下：1.權威性方面：提供網站負責人資訊及連絡方式、資料的新穎性；2.網站贊助者方面：網站贊助者的相關資訊；3.內容方面：清楚陳述目的、說明適用的對象、內容對適用對象而言是否得當、網頁中沒有拼音或文法上的錯誤、說明文獻來源、遵守版權法相關規定、串連是否有效、內容正確無偏見；4.影像方面：影像的數量足夠、影像與網頁有所關連、影像傳輸速度快、影像是否為 GIF 檔；5.呈現方式方面：容易閱讀、格式一致、網頁上有足夠的空白、多重網頁取代單一長網頁、下載時間合宜；6.多重網頁使用方面：母頁簡短且包含網站目的、內部網頁能串連到母頁、較長的網頁能提供內容串連；7.加強整體網頁方面：含蓋圖形、影音及影像資料；8.技術上的考量方面：所有的串連都是有效的、每頁最後提供有效的 e-mail 連結、計時器有效運作、提供影音輔助所需的設備及軟體。

Tate & Alexander (1996) 提出的網站查核標準及項目如下：1.權威性方面：說明贊助者、提供資料檢驗贊助者的合法性、提供贊助者的聯絡地址和電話、說明網頁製作者及其背景資料、說明版權擁有者；2.正確性方面、資料的真實性、沒有語法、拼音和印刷上的錯誤、說明負責資料正確性者、統計資料以圖表清楚呈現；3.客觀性方面：提供的資訊為公用、廣告與資訊內容有明顯區別；4.新穎性方面：說明網頁編寫及上網、最近更新之時間、提供訊息以證明資料的新穎性、以圖表清楚的呈現資料、若資料以不同版本呈現，應說明此網頁屬於那一版本；5.含蓋範圍方面：說明網頁是建構中或已完成、說明是否有與網頁上等同的紙本資料。

Druse (1991) 提出評鑑網站的標準包括以下幾部分：1.目的 (purpose)：清楚的陳述目的、有預設的使用族群；2.權威性 (authority)：網站的管理機構、提供聯絡方式、網站上所有的資訊具有版權；3.內容 (content)：資訊的正確性、時

效性、一手資料或二手資料、是否有偏見、可靠性、主題是否詳細、提供超連結之敘述性說明，是否有圖形或影像以加強內容、是否提供廣告；4.技術上的品質（technical quality）：便利使用者、容易航行、首頁簡明、快速瀏覽、網站呈現的架構容易幫助使用者找到資訊、內容互動性、網站穩定、提供搜尋引擎等；5.評論（review）：被專業機構評比過。

翁慧娟（民 88，頁 172-173）對國內網站品質進行評鑑研究，研究結果建議評鑑大學圖書館網站可採用下列的標準：連結性、畫面設計、資源分類架構、內容、權威性、選擇性、新穎性、多元性、管理資訊、服務溝通、溝通性、及使用者調查等。

綜合文獻資料，發現網路資源的系統設計者、中介者和使用者對評鑑指標有不同的看法。系統設計者的評鑑標準較多，包括下列各項：技術新穎性、個人化服務、提供檢索查詢功能、多媒體的呈現方式、是否提供相關網站連結、即時的互動性線上訂閱、儲存功能、互動速度、多媒體的呈現、即時的互動、專人負責、內容是否適合以此媒體呈現、資訊豐富、資訊內容合法性、分類明確、親和的使用者介面、層次架構清楚、架構完整性、網站有無使用資格限制、網站簡介、網頁與網頁間的連結性、是否滿足使用族群的需求、下載速度、系統安全性、系統穩定、適用任何瀏覽工具、連線速度、網頁更新頻率、網頁畫面與主題的相關性、資訊正確性、內容更新速度、畫面應美觀與簡潔、獨特性、資訊實用性、權威性、版面配置適當、主題範圍、內容深度、及提供新知等。

中介者的評鑑標準可分為十五項屬性，分別是：資源分類架構、畫面設計、新穎性、檢索功能、輔助說明、索引功能、資訊內容、互動性、連結功能、整體設計、權威性、可讀性、獨特性、網站管理資訊、回應時間等。使用者的評鑑標準則是：新穎性、資料內容、畫面設計、使用者介面、速度、互動性、網站管理資訊、航行功能等。

整合設計者和中介者的意見，評鑑項目應包括：資訊內容、美工設計、使用性、互動性、連結性、搜尋及檢索功能、人機介面、資源分類架構、輔助說明、整體設計、網站管理資訊、回應時間等，這些評鑑項目已可發展為基本的評鑑查核表。然而，網站的查核標準和查核項目仍應依評鑑對象和評鑑目的不同而有不同的內容和設計，評鑑研究的目的才能準確達成。

四、數位圖書館

數位圖書館的評鑑研究深受各數位圖書館計畫和個別研究者的關心(如 Belkin, 1994；Bishop, 1995；Mead & Gay, 1995；Goral, 1998；Druse, 1999；張嘉彬, 1999

等)。張嘉彬(民 88, 頁 35)整理幾位研究者提出來的線上資訊檢索系統和搜尋引擎的評鑑指標,發現將「資訊內容」列為評鑑標準有六位研究者、「權威性」有五位、「檢索功能」、「系統管理資訊」及「航行功能」等三項指標有四個研究者、將「整體架構及組織」、「檢索結果展示」、「連結性」、「互動性」為評鑑指標有三個研究者;「輸出控制」、「輔助說明」、「目的」、「畫面設計」各有二個研究者、「檢索表現」、「技術上的考量」、「社會層次的影響」等三項評鑑指標各有一研究者選為評鑑指標。

綜合上述的討論,數位圖書館基本的評鑑指標可有以下十四項:網站管理資訊、畫面設計、航行功能、連結性、資訊內容、資源分類架構、檢索功能、檢索結果展示、新穎性、互動性、可讀性、輔助說明、使用者調查、及特殊功能等。其他評鑑指標可依評鑑目的和評鑑對象適度選用。

肆、數位圖書館 / 博物館評鑑的議題

除了評鑑指標之外,從事數位圖書館 / 博物館評鑑仍要了解以下幾個重要的議題。

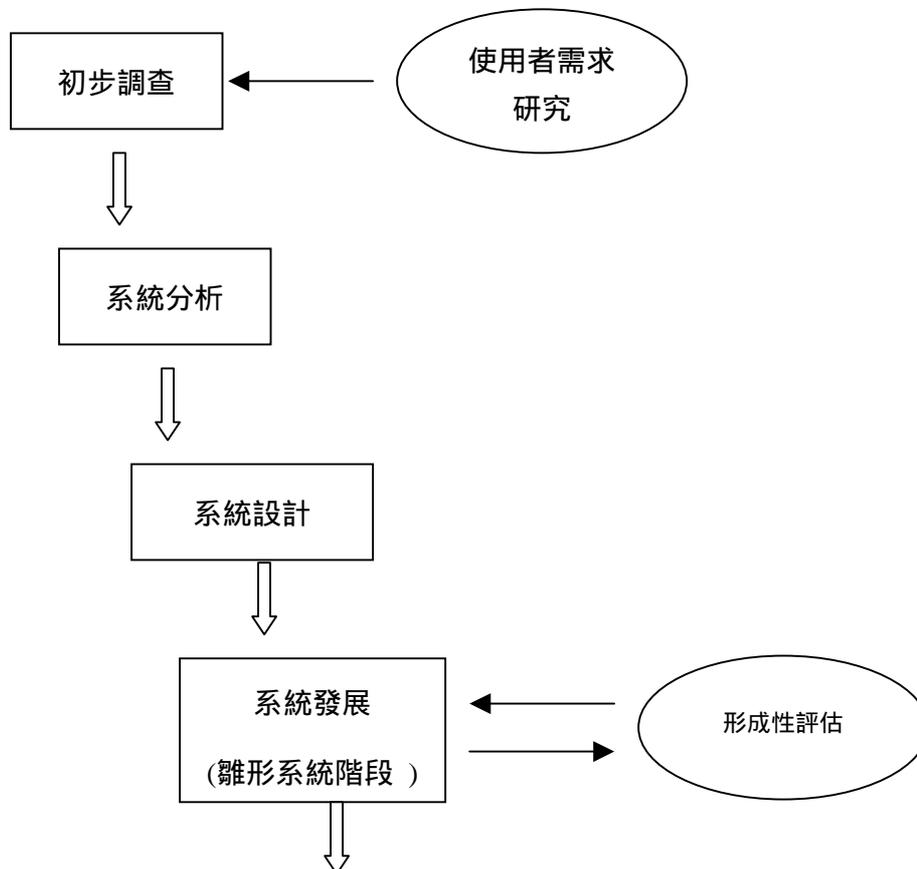
一、了解系統評鑑的層次

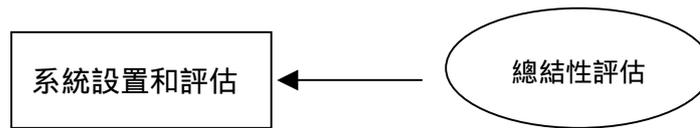
「美國數位圖書館先導計畫」(The DLI)有一非正式研究小組,每一年舉行兩次工作報告。在 1995 年的春季會議中,該小組討論以使用者為中心,將評鑑層次歸為四方面:數位館藏的合適性、系統功能、界面及使用性評估;檢索表現及使用者行為評估;系統評鑑對於使用者工作改變的影響;系統評鑑對於公共政策的影響(Bishop, 1995)。Saracevic(1995)認為系統評鑑依目的不同有六個層次:工程層次(engineering level)、輸入層次(input level)、處理層次(processing level)、輸出層次(output level)、使用與使用者層次(use and user level)以及社會層次(social level),其中:(一)工程層次(engineering level),譬如軟體和硬體的功能,包括系統的穩定程度、錯誤率、速度、維護和彈性等,資料檢索方法和計算式(algorithms)的運算之效力(effectiveness)和效率(efficiency);(二)輸入層次(input level),輸入了什麼資料,內容為何,資料庫的範圍等;(三)資料處理層次(processing level),資料處理的技術、計算式以及方法等;(四)輸出層次(output level):檢索的過程、和系統的互動、系統回饋,以及獲得的檢索結果等;(五)使用和使用者的層次(use and user level):使用者的檢索問題和採用的檢索策略、市場行銷、系統的合用等;(六)社會層次(social level):資訊系統對社會的影響,譬如對某個領域之研究發展、生產力、決策過程等產生的影響。層次一至四是系統功能導向的評鑑,層次五至六為使用者中心的使用者評鑑。評鑑者應確認系統評鑑的目的,視實際需要決定系統評鑑的層次(吳美美, 民 87)。

二、了解系統評鑑的適當時機

依數位圖書館/博物館發展階段及評鑑目的，可將系統評鑑分為「需求評估」、「形成性評估」與「總結性評估」。在評鑑計畫一開始時，研究者或系統開發者應做初步調查，調查使用者需求並進行分析；而在雛形系統的建置過程中，系統設計者應進行形成性評鑑，了解系統目前的問題、優缺點，作為持續改進的參考。系統開發完成，系統所進行的則是更全面的總結性評鑑。

一般的系統發展程序可分成需求調查分析、系統分析、系統設計、系統發展、系統設置和系統評鑑。根據不同的發展階段，系統管理者應採用不同的評估方式。在系統開發的初步調查階段，系統建置者必須了解使用者的需求，並據此進行系統分析與系統設計；系統規畫完成，且發展出一套雛形系統時，系統便應著手進行形成性評鑑，並視需要持續進行，換言之，形成性評估不應只是在雛形系統完成後所進行的一次評估，而是在系統不斷發展與繼續更新的過程中持續進行的評估活動；至於總結性評估，則是一旦系統開發完成後，評估者所進行全面性的完整評估研究。系統開發階段與評鑑實施的關係可由圖一表示。





圖一 評鑑研究與系統開發

三、 決定系統評鑑的目的

有關「系統評鑑」，我們常聽到幾個名詞：系統評鑑、使用者評鑑、使用者研究、系統評估、形成性評鑑、總結評鑑等，這幾個名詞之間確有密切的關係。早在 1950、60 年代，資訊檢索評鑑研究在實驗室進行的階段，稱為系統評鑑(system evaluation)，到了 70 年代，市面上已經可見提供使用者使用的檢索系統(operational system)，研究者便進行使用者評鑑(user evaluation)以了解檢索系統的成效。前者注重系統內部功能的良窳，後者注重使用者滿意程度，包括檢索問題是否獲得合適的檢索結果，以及人機互動是否友善易用等(吳美美，民 87；Wu & Sonnenwald, 1999)。使用者研究(user study)是以使用者和使用者的資訊需求為主，通常有兩種用途，其一，作為系統設計之前的需求分析，在進行此種目的使用者研究時，通常脫離特定的檢索系統，純粹為了解使用者的資訊行為為主，使用者需求分析是系統設計之前需要了解的重要資料；其二，為了解系統的好用與否，獲得的研究結果主要用以決定系統的滿意程度。此外，資源管理、系統的效益和投資報酬都是系統經營者關心的課題，評鑑研究依評鑑目的的不同，應有不同的設計，如研究方法和評鑑指標的決定等。

四、 依系統評鑑的目的決定評鑑指標

在資訊檢索的評鑑研究中，較常使用的評鑑標準是檢索效率 (efficiency) 和檢索效益 (effectiveness)，而精確率 (precision) 和回現率 (recall) 更是系統評鑑常用的指標，但是由於這兩個評鑑指標只能用於特定的評鑑目的，如檢索技術的層面等，並未考慮使用性(usability)和使用者滿意程度(goal satisfaction) (Smeaton & Harman, 1997, p.170)，更未考慮更高階層的評鑑目的，例如系統經營管理或對機構、社群、人群的影響，因此評鑑研究中最需思考的仍是如何依不同的評鑑目的找出適合的評鑑指標，以下分五個層次簡單討論。

(一) 依管理的需要

一般而言，評鑑指標依評鑑目的不同而有不同，以系統開發單位而言，在管理方面，由於投資效益和使用推廣是關心的目標，可考慮利用 Rogers(Wu & Sonnenwald, 1999)創新傳播的五個特性：相對優勢、適合性、複雜性、可試性、可觀察性等作為評鑑標準，而評鑑指標可考慮使用：系統相關、主題相關、速度、

經濟效益、動機相關、組織相關、社會相關、使用容易度、認知相關、情境相關、易於試驗、易於示範的程度等（如表二）。

表二 管理者系統評鑑指標（修改自 Wu & Sonnenwald, 1999）

評鑑標準	評鑑指標	評鑑項目
相對優勢	主題相關程度 需求相關程度 速度 經濟效益 便利性	回現率及精確率 資料內容深度與廣度 系統回應時間 費用效益分析 是否普及 易於接觸
適合性	和個人動機的相關程度 和機構需求的相關程度 和社會需求的相關程度	符合使用者的期望 符合組織的期望 符合社會的期望，迎合公共政策
複雜性	使用容易程度 認知相關程度 情境相關程度	任務完成時間、容錯誤率 使用者對於系統操作接受程度 使用者問題或工作解決滿意程度
可試性	可試驗性(ease of experimentation)	可取用、訓練時間長短、訓練費用高低
可觀察	可示範並顯示效果	使用者滿意程度

（二）為了解使用者資訊需求

系統以人為中心，最終為滿足人的需求。因此，系統評鑑的目標之一為找出各種使用者的需求，包括已知使用者、潛在使用者、與待開發之使用者。評鑑指標，包括研究使用者是那些人？他們的需求是什麼？這些問題可透過使用者評鑑研究獲得進一步的了解，充實數位圖書館的內容。

（三）為改進系統功能

為改進系統的功能，需要了解使用者的特質、資源的特質、和系統的原有功能；亦即了解使用者的使用（usability）和接受程度（user acceptance），例如使用者的使用率、滿意程度等。相關的評鑑問題包括此數位圖書館有那些資源？使用的情形為何？支援生活或工作的那些層面？系統介面設計是否易用等。了解系統所涵蓋的資源特質和範圍、系統是否為使用者接受等，系統功能才能獲得改進。

（四）為了解使用者和系統互動

數位圖書館的評鑑研究可以涵蓋很多層面，但使用者互動評鑑研究卻是其中十分重要的項目。數位圖書館的建置主要在協助人們尋找資訊，雖然更賦有典藏

資源的重任，但使用者是否從系統互動中學習新知、找到所需、豐富愉快的使用經驗仍是數位圖書館能否永續發展的重要關鍵，且是系統改進建議的來源，互動研究除靠系統內部磁帶紀錄，更需靠觀察法和訪談法等耐心收集資料、分析資料，研究結果可做為系統介面設計參考。

(五) 為了解系統對人類社會的影響，社會關懷層次

作為社會資源的數位圖書館/博物館資訊系統，若以一般社會人士或中小學師生為主要服務對象，所要評鑑的便是「這樣一個系統對社會有什麼影響？對教育活動有什麼影響？如何影響？」評鑑結果，可以做為是否繼續資助、推動類似系統進一步發展的決策依據。

五、評鑑的方法

數位圖書館的評鑑方法傾向多元研究方法，包括小樣本的質性研究、大規模的使用者調查，訪談、觀察、查詢記錄分析等都是常用的評鑑研究方法。以下簡介幾個數位圖書館評鑑計畫所採用的研究方法。

加州大學聖塔芭芭拉分校「亞歷山卓計畫」(Alexandria Digital Library Project)，主要的目的在發展以使用者為主、具親和性的，同時能提供地圖、影像和其他特殊資源的整理、瀏覽和檢索之綜合性系統，希望能解決傳統圖書館文件和特殊檔案分開處理的問題。為評鑑系統效益，該計畫成立「使用者評鑑團隊」(user evaluation team)，希望提供使用者介面改善之建議。除了使用語言分析、分析使用者的查詢記錄，比較使用者載入的系統版本等方法外，並採用「說出想法」法(think-aloud)，但發現效果並不好，因為使用者在使用介面時並不習慣說話，而說話時就停止使用系統。因此研究人員改用錄影的方法，讓使用者在看自己的錄影帶時再予以說明。蒐集資料的工具包括：互動查詢記錄、錄影帶、問卷、「說出想法」所彙整之資料(Buttenfield, 1995)。

加州大學柏克萊分校「環境規畫與地理資訊系統研究」，所採用的方法有：深度半結構化訪談、焦點團體、語言分析(protocol analysis)、使用者行為的觀察、文獻分析等方法(Van House, 1995)。

「美國記憶」數位圖書館計畫，其評鑑研究收集的資料包括量化資料有1801份問卷、44個機構中收到21個機構的查詢記錄；質化資料則有55位職員及121位使用者的個別訪談、電話訪談、資料庫瀏覽(觀察)。「美國建置數位圖書館計畫」的評鑑研究中，採用的評鑑方法包括面談、開放式問卷、焦點團體等方式(Kilker & Gay, 1998, p.63)。「美國政府資訊指引服務」評鑑計畫，所採用的評鑑方法包括資料庫瀏覽(site visits)、焦點團體(focus groups)、調查法(survey)、美國政府

資訊指引服務(GILS)記錄內容分析 (GILS record content analysis)、線上使用者評鑑 (scripted online user assessments)、查詢記錄分析 (web server transaction log analysis)、政策及文獻探討 (policy and literature review) (Moen & McClure, 1997)、張嘉彬評鑑台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統，採用文獻分析法、問卷法、查詢過程記錄分析法、不介入觀察法及深度訪談法等。

由以上各個數位圖書館評鑑計畫所採用的方法來看，質化研究方法是一種趨勢。質性研究具有下列的特色：開放、有彈性、完整、個案導向、歸納和非控制性等。這些特質使得質性研究，是找出人類內在行為的最佳方法。

量化分析是實驗室系統評鑑的主要評鑑方法，但其缺點除了評鑑指標缺乏彈性外，通常還有：資訊需求被事先定義，且固定不動、使用者真正的需求很少反應在研究中、專業人員和使用者的相關判斷很少被比較、常使用主題相關判斷卻忽略相關是多面向的概念，以及忽略情境相關的重要 (Fidel, 1993, p.226)。

雖然質化研究方法愈來愈為研究者重視，但是未來的數位圖書館評鑑研究仍要兼顧實驗室和實際使用者兩方面的研究需要，以評鑑研究目的為導向來決定評鑑研究的方法才是正途。同時採用質化和量化研究方法，採用多重研究方法，將是研究趨勢，採用多元研究方法蒐集獲得的資料及分析結果，可互相對照，研究結果可望更為完整、可靠。

伍、結論

數位圖書館的建置需要各方面的專業人員合作努力，結合圖書館、資訊檢索、和電腦多媒體設計等專業人材，以及使用者共同參與。透過評鑑，不僅可以了解系統的功能及使用情況，並能了解使用者行為，可做為改進系統之參考，增加使用率。數位圖書館在開發至建置完成，需有一套標準的評鑑流程，在不同的開發階段予以檢驗，俾達到最佳效益，使其得以永續經營，受使用者歡迎。

評鑑研究是數位圖書館建置過程中一個不可或缺的部分，使用的方法也傾向多元。除了傳統量化分析方法之外，質性研究已成為趨勢之一，依評鑑目的善加選用合適的方法。而評鑑的角度也已由系統功能的表現，轉向使用者需求的滿足。由使用者的角度來評鑑數位圖書館的狀況，其中要考量的層面更廣，除了使用者的需求外，還包括影響需求的因素，如使用者的環境、經驗、社會背景等，都值得系統改進參考。評鑑的指標隨評鑑的目的而有不同，建立一套合用的數位圖書館評鑑指標是必要的。

每個數位圖書館計畫都需要隨著系統開發的不同階段，給予不同性質的評鑑

研究。一份評鑑手冊提供評鑑指引是必要的。手冊的內容除了應提示系統在不同開發階段所應進行的不同評鑑研究外，評鑑研究的要領、方法、程序，如評鑑的目的、評鑑的時機、評鑑的方法、評鑑指標、評鑑研究的流程、注意事項、評鑑報告的撰寫方法等都要詳細陳述。國科會推動數位圖書館博物館計畫，不能省略評鑑研究這個重要的部分。黃慕萱教授領導一個數位圖書館評鑑研究計畫(黃慕萱、林珊如、吳美美，民 88)，以淡水河溯源為例，便是數位圖書館評鑑研究的一個範例。但是一個具參考性質的評鑑手冊仍有待產生，其他數位圖書館/博物館計畫才有具體的評鑑方法可資依循。我們期待這個數位圖書館的評鑑研究標準流程趕快建立。

誌謝：感謝國科會贊助本研究計畫「數位博物館專案先導計畫 - 系統評估規範」(計畫編號：NSC 88-2745-P002-013)，本文為文獻探討之部分結果。淡江教資所姜杏蓉同學不厭其煩，協助繕稿，特一併致以誠摯謝忱。

參考文獻

- 吳美美 (民 87)。「邏輯相關」和「適切相關」--中文資訊檢索系統評鑑量標初探。圖書學與資訊科學 24 卷 2 期，頁 44-64。
- 黃慕萱、林珊如、吳美美(民 88)。數位博物館專案先導計畫--系統評估規範之建置自評報告。(未出版)。
- 張嘉彬 (民 88)。台大電子圖書館與博物館 - 臺灣平埔族探源雛形系統之評鑑研究。國立臺灣大學碩士論文。
- 翁慧娟 (民 88)。我國大學圖書館網站品質評鑑之研究。國立臺灣大學碩士論文。
- Belkin, Nick. (1994). Digital libraries: interaction and evaluation. Talk given in the Information Retrieval as Interaction Workshop. Academia Sinica, Taipei, August 30-31, 1994.
- Bishop, Ann Peterson. (1995). Working towards an understanding of digital library use: A report on the user research efforts of the NSF/ARPA/NASA DLI projects. D-Lib Magazine. (<http://www.dlib.org/october95/10bishop.html>)
- Borlund, Pia., & Ingwersen, Peter. (1997). The development of a method for the evaluation of interactive information retrieval systems. Journal of Documentation, 53, 225-250.
- Buttenfield, Barbara P(1995). User evaluation for the Alexandria digital library project. (<http://edfu.lis.uiuc.edu/allerton/95/s2/buttenfield.html>)
- Courtois, Martin P., Baer, William M., & Stark, Marcella. (1995). Cool tools for searching the web : A performance evaluation. Online, 15, 15-32.
- Covi, Lisa M. (1995). Evaluation of digital library use: How can we understand how contributors create and use scholarly communication ? (<http://edfu.lis.uiuc.edu/allerton/95/s5/covi.html>)

- Dong, Xiaoying, & Su, Louise T. (1997). Search engines on the world wide web and information retrieval from the Internet: A review and evaluation. Online & Cdrom Review, 21, 67-81.
- Druse, Judy. (1991). Web site evaluation criteria.
(<http://www.washburn.edu/mabee/reviews.html>)
- Fidel, R. (1993). Qualitative methods in information retrieval research. Library and Information Science Research, 15, 219-247.
- Frew, J., Freeston, M., Kemp, R. B., Simpson, J., Smith, T., Wells, A., & Zheng, Q. (1996). The Alexandria digital library tested. D-Lib Magazine.
(<http://www.dlib.org/october95/10bishop.html>)
- Goral, Ron. (1998). Web site evaluation checklist.
(<http://www.madison.k12.wi.us/webpub/evaluate.htm>)
- Janes, J. (1995). <http://edfu.lis.uiuc.edu/allerton/95/s1/janes.html>
- Kelleher, J., Sommerlad, E. & Stern, E. (1996). Evaluation of the electronic libraries programme: Guidelines for ELIB project evaluation.
(<http://www.ukoln.ac.uk/services/elib/papers/tavistock/evaluation-guide/>)
- Kilker, Julian, & Gay, Geri (1998). The social construction of a digital library: A case study examining implications for evaluation. Information Technology and Libraries, 17, 60-70.
- Lancaster, F.W., Humphrey, S. M., Ulvila, J. W., Smith, L. C., Allen, B., & Herner S. (1996). Evaluation of interactive knowledge-based systems: overview and design for empirical testing. Journal of the American Society for Information Science, 47, 57-69.
- The Library of Congress. (1993). Final report of the American memory user evaluation, 1991-1993. (<http://lcweb2.loc.gov/ammem/userval.html>)
- McMurdo, George. (1998). Evaluating web information and design. Journal of Information Science, 24(3), 192-204.
- Mead, June P., & Gay, Geri. (1995). Concept mapping: An innovative approach to digital library design and evaluation.
(<http://edfu.lis.uiuc.edu/allerton/95/s2/mead/mead.html>)
- Moen, William E., & McClure, Charles. (1997). An evaluation of the federal government implementation of the government information locator service (GILS): Final report. (<http://www.unt.edu/slis/research/gilseval/>)
- Pejtersen, A.M., Dunlop, M. & Fidel, R. (1999). A user centered framework for evaluation of the web. ACM SIGIR '99 workshop on evaluation of web document retrieval. Berkeley, C.A., USA August 19, 1999.
- Polly, Jean Armour, & Lyon, Elaine (1995). Out of the archives and into the streets: American memory in American libraries. Online, 15, 51-57.
- Saracevic, T. (1995). Evaluation of evaluation in information retrieval. *Proceedings of*

- the 18th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp138-146.) New York: ACM.
- Smeaton, Alan F., & Harman, Donna (1997). The TREC experiments and their impact on Europe. Journal of Information Science, 23, 169-174.
- Su, Louise T. (1992). Evaluation measures for interactive information retrieval. Information Processing & Management, 28, 503-516.
- Tate, Marsha, & Alexander, Jan. (1996). Teaching critical evaluation skills for world wide web resources. Computers in Libraries, 16(10), 49-55.
- Van House, Nancy (1995). User needs assessment and evaluation for the UC Berkeley electronic environmental library project: a preliminary report.
(<http://csdl.tamu.edu/DL95/papers/vanhouse/vanhouse.html>)
- Wu, Mei-Mei & Sonnenwald, Diane (1999). Reflection on Information Retrieval Evaluation. In: Proceedings of 1999 EBTI, ECAI, SEER, & PNC Joint Meeting, Taipei, Academia Sinica, January 18-21,1999, pp.63-81.