

網路資源備份保存技術與應用探析

洪淑芬

摘要

本文主要在探討網路資源備份保存之必要性；透過國外網路資源保存之實例，介紹相關技術實務與合作規劃，並說明同時保存網路上「開放式資源」與「閉鎖式資源」之重要性。進而實際使用國外保存網路資源機構 Internet Archive 之資料庫，以測試臺灣各類機構網路資源備份保存之狀況，藉以了解網路資源本身結構對於完整備份保存之影響。網路資源製作上的規劃，與其是否得以被完整備份保存，關係密切。而網路資源本身之 Metadata 又將影響自動化編目的可行性與正確性，而間接影響備份保存之資料的提供利用。因此，本文試圖涵蓋網路資源保存上必須列入考量的因素，包括網路資源建置端，以及專為備份保存網路資源而設立之網站機構端雙方所需的努力；期能喚起注意、引發相關計畫之執行，並促成臺灣

關鍵詞 (Keywords)：網路資源備份保存；網路資源搜尋；網路資源可即性；網路開放性資源；網路閉鎖性資源；機構典藏；全域模式；選擇性模式；主題式模式；寄存式模式；混合式模式；網頁可即性檢核工具；網頁可即性輔助工具

Web Preservation；Web Archiving；MINERVA；PANDORA；Internet Archive；Kulturarw3；WayBack Machine；Web Clawing；Open Access Resources；Close Access Resources；Institutional Repository；Whole Domain Model；Comprehensive Approach；Selective Model；Thematic Model；Deposit Model；Combined Approach；Web Information Accessibility；Accessibility Verifying Tool；Accessibility Software

洪淑芬：國立臺灣大學圖書館特藏組編審、國立臺灣大學圖書資訊學研究所博士生；E-mail: shufen@ntu.edu.tw

網路資源本身結構之改善，裨益數位文化資產之保存以及其長久可即性。

一、前言

臺灣近幾年來網路資訊環境發展快速，而且政府亦十分重視網路資訊內涵之提升。因此，除了國家資訊基礎建設之硬體環境創造外，亦持續投注經費於文化資產之數位化。另一方面，在資訊網路化與電子期刊漲價的時勢之下，非屬於私人出版社之數位出版，例如由學術社群所組成的網路出版，成為新時代的產物。網路的內容不斷地豐富化，各類網頁的數量也持續激增，如何得以保存維護實體文化資產一般地保存網路上的資源，成為迫切需要考慮的問題。目前國外先進國家不但已經體認到網路資源保存、以及有系統地進行系列性備份之重要性，也實際推動搜尋保存之計畫，進行網路資源之備份存檔，並進行編目整理，期許為現代人提供「全人類的知識」^[1]，以及將這一代的文化資產提供給未來的世代。^[2]

反觀臺灣政府機構及學界對於網路數位資源之創造與整理，目前尚止於建置入口網站（Portal）以方便使用者能夠充分瀏覽相關資源、以及達成跨資料庫檢索之目標而已。例如：國家圖書館的「臺灣研究入口網」（<http://twstudy.ncl.edu.tw/>）與「華文知識入口網」（<http://search.ncl.edu.tw/LISP/hypage.cgi?HYPAGE=home.htm>）、數位典藏國家型科技計畫的「數位典藏聯合目錄」（<http://catalog.ndap.org.tw/dacs5/System/Catalog/Catalog.jsp>），行政院文化建設委員會「國家文化資料庫」（<http://nrch.cca.gov.tw/ccahome/index.jsp>）、以及圖書館機構為讀者整理的各類「網路資源」^[3]等，都屬於入口網站，並無法達成網路資源永久有效保存之目標。有鑑於現今網路上的豐富資源與本世代之網路文化，無論對於當下或未來，都具有保存與研究上之價值，而臺灣對於網路數位資源保存之工作尚付諸闕如，本文乃試著從資料保存、前瞻性的考量、未來資訊可即性的穩定性等方面，比較「入口網站」與「網路資源備份保存網站」二者所能夠發揮的功能之不同，剖析並倡議

[1] Internet Archive 首頁揭示出其理想。“Universal Access to Human Knowledge,” Retrieved June 1, 2006, from <http://web.archive.org/index.php>.

[2] Library of Congress, “About MINERVA,” Retrieved June 1, 2006, from <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-about.html>.

[3] 例如香港科技大學圖書館由館員整理出的“Internet Resources,” Retrieved June 1, 2006, from <http://library.ust.hk/res/beyond/>。該網站除了提供關鍵字與片語的檢索之外，亦提供主題分類以及依網站字母順序排列的清單，方便查詢使用，堪稱網路資源入口網站之範例。

網路資源備份保存之必要性，期能喚起國內圖書與資訊界專家學者、以及決定國家政策的相關單位與決策者，共同關注此議題。

網路資源備份保存之實例，對於我國推動相關計畫，必然有所助益。因此，本文乃提出國外較具代表性的成功實例，介紹網路資源保存之相關規劃與實務技術。另外，經備份保存後的網路資源，必須經過良好的組織，方能即時或在未來提供使用者檢索利用。然而因為網路資源數量龐大，極大部分的組織工作有賴於自動或半自動的方式進行組織。要達到網路資源之自動組織，數位資源本身 Metadata 之健全性是主要關鍵。另一方面，在網路資源之備份過程中，網頁與網站本身的結構亦將影響備份之完整性。換言之，事後的保存（蒐集已產出且存在於網路上的資源），必須有事先確實嵌入的數位資源 Metadata 之支援，以及良好的網頁結構，方可順利達成數位資源之組織以及完整備份保存。美國的網際網路協會（World Wide Web Consortium，簡稱 W3C）早已注意到此一重要議題，故制定「網頁內容可即性指導方針」（Web Content Accessibility Guidelines）^[4]，研擬網站與網頁的評核要點，並有多家軟體公司根據該指導方針，開發設計可以幫助數位網站資源在設計與結構方面符合標準之軟體。

數位資源本身的健全性，以及備份保存機構之永久備份典藏，二者之間是共生共榮的關係。機構本身建置結構與資訊完善的網路資源的好處，是未來可以在網路資源備份保存網站之檔案櫃中，查詢或存取任一時期的舊網頁資料，亦即在獲得免費的備份保存服務之餘，又可存取過去完整的歷史網頁。由於本文之意旨在促成國內成功地推動網路數位資源備份保存計畫，故對於具有連帶性關係之網站資料編目議題、以及如何使機構所建置之網站資料能被完整備份存檔，亦一併探討之。除了期望喚起國內學者及決策者注意進而推動相關計畫之外，亦能幫助臺灣數位資源產出機構、網路資源備份保存機構二者之間，建立良好的互相依存關係；使臺灣未來的網路資源備份保存，由各界共同基於同樣的目標，共同進行必要的完善規劃，無所偏廢，而能獲致最佳備份保存效果。

二、何謂「網路資源備份保存」

本文所討論的「網路資源備份保存」是譯自英文的 Web Preservation 與 Web

[4] W3C, *Web Content Accessibility Guidelines 1.0*. (1999), Retrieved July 16, 2006 from <http://www.w3.org/TR/WAI-WEBCONTENT/>.

Archiving。在數位資源快速增加的現實環境下，典藏機構早已意識到數位資源保存的重要性。但是，國內過去幾年大力推行政獻、文物典藏數位化計畫，所重視的數位資源保存大致是指數位化成果之維護保存，推廣利用方面則是透過單一或少數「入口網站」。相對於「入口網站」的資料組織與保存方式，另外有一種網路資源之保存是將網路資源加以備份，並進行必要的整理編目，而且甚多情形是將整個網站加以重複備份保存；此種網路資源之保存稱為 Web Preservation 或 Web Archiving。美國較具規模地進行網路資源保存的機構，包括美國國會圖書館、Internet Archive 等。美國國會圖書館成立網路資源保存計畫任務小組（LC Web Preservation Project Team），其任務是執行網路資源保存計畫（Web Preservation Project），該計畫負責網路資源之選擇、收集、編目、促成已備份存檔之網站的可存取性。^[5]而 Internet Archive 之網站雖然進行廣泛收集而未強調選擇，但亦進行編目與提供檢索利用，本文第五節有進一步之說明。此外，更早執行此類計畫的澳洲 PANDORA 計畫，於討論相關計畫措施時，則採 Web Archiving 之詞。^[6]換言之，各者雖然範圍（Scope）有別、名稱有異（有的採用 Web Preservation；有的採用 Web Archiving），然而皆肩負定期將網路資源存檔備份、加以編目，以提供使用者查詢過去之網站或網頁資訊的使命。因此就此意義而言，Web Preservation 與 Web Archiving 二詞可互通（Interchangeable）。為了與「入口網站」之資料整理有所區別，本文乃將 Web Preservation 與 Web Archiving 譯為「網路資源備份保存」，以真實反映實務上備份、存檔保存的作法。

「網路資源備份保存網站」與臺灣目前許多機構建置之「入口網站」，二者之目的與功能不同：「入口網站」所提供的資料如經更新，使用者即無法再查得舊有的資料；舊有資料的存檔備份是各機構之責。而「網路資源備份保存網站」則依據所訂定的規則，定期或不定期地蒐集網路資源，進行備份保存。而對於網路資源之收集，亦訂有搜尋與收集保存政策。被列為其收集對象之網站資料，一旦經備份保存，將永遠存留，使用者可以查詢網站中過去曾經提供之資料。

[5] Arms, B. and Members of the LC Web Preservation Project Team, "LC's Web Preservation Project: a Pilot for Selecting, Collecting, Cataloging, and Accessing Archived Websites: PPT, at Mumford Room, Madison Building," (2001), Retrieved July 16, 2006 from <http://www.loc.gov/minerva/presentations/minervafeb2001.ppt>.

[6] National Library of Australia, "Web Archiving," Retrieved June 1, 2006, from <http://www.nla.gov.au/padi/topics/92.html>.

「網路資源備份保存網站」有助於國家網路資源之永久性典藏，且是支援「入口網站」的重要資料來源，因此是先進國家極為重視之議題。以下茲以表一進行「入口網站」與「網路資源備份保存網站」之性質與功能比較，藉以呈現二者功能面、服務面、資源保存面等之差異，以利討論何以國內需考量建置國家的網路資源備份保存網站，並提供規劃上的參考。

表一：「入口網站」與「網路資源備份保存網站」之比較

比較項目		入口網站 Portal	網路資源備份保存網站 Web Preservation / WebArchive Website
目的	資料保存	著眼於所擁有的以及可掌控之資料庫的保存。	著眼於網路文化資源之典藏保存。
	資料提供	1.整理現存網路資源，提供大量現存網路資源之連結。 2.彙集主題性資料庫，方便查詢利用。	備份保存網路資源，以自動、半自動、或人工整理已備份之網站資源，提供查詢舊資料。
	前瞻性的考量	著眼於當下的服務。	著眼於服務當下與未來的世代。
開放考量	可見度	為機構網站中明顯重要之項目，可見度高。	可能以獨立網站之型式存在，可見度較低；亦即民眾學者對其熟悉度較低。
	著作權	可公開流傳。	所蒐集保存之網路資源，如屬個人創作或擁有著作權之資料，經網路資源備份保存網站加以整理編目後，成為具有「獨立出版品」一般的性質，則透過網路再提供使用，需審慎評估其可否，並尋求解決著作權問題之道。
查詢服務	「資料可即性」的穩定性	網路上的網站消失，資訊即不可得。	網路上的網站消失，舊有資訊仍可查得。
	可查尋的深度	可深入資料庫中查詢。	如果資料庫設有權限管制，即無法進入蒐集備份保存，而無法提供資料庫的查詢。
人力	能力需求	1.主要需求為網站與資料庫建置能力。 2.資料庫內的書目資料為人工建置。	1.需要的資訊技術能力包括網站與資料庫建置、自動網路資源搜尋備份、自動資料分析與分級提供使用之流程設計等技術。 2.自動編目之技術、部分搭配人工進行檢核。
	維護需求	需時時檢驗網站是否存在、連結是否有效。	一旦保存，無需檢驗連結的有效性；存檔當時的連結狀況無法改變、也不試圖去改變。
	備份方式	人工處理較多。	相當大的部分可自動化處理。
硬體需求	伺服器與儲存空間	所需儲存空間較小。	所需儲存空間大，而且成長快速。

資料來源：本文著者研究心得。

三、網路資源備份保存的必要性

網路資源保存方式不同於實體物件之典藏，其既具有易於多重備份之特性，卻又極易於不經意之間瞬時消失。臺灣近幾年來大力推動數位典藏計畫，也重視數位化成果之保存。但是對於數位化成果之保存，僅限於利用各種儲存媒體、磁碟陣列、電腦主機等備份；而對於數位化成果之推廣利用，則是透過典藏機構本身的網站，以及「數位典藏聯合目錄」或「國家文化資料庫」之類的入口網站提供使用，亦即大致是以單一或雙管道提供由資料文物典藏機構數位化完成之資料。對於一般的網路資源，包括各種開放式或閉鎖式的網路數位資源，則未討論到其備份保存的課題。^[7]但是基於下列理由，有系統地定期備份保存網路資源極為重要。

(一)永久典藏「唯數位版產出」(Born Digital)之資料

基於業務推廣目的，各種公、私立機構均在其網站上張貼各種與業務相關之資料，而此類資料隨時可能因業務內容、相關規則之變更而變更。此外個人亦會建置個人網站。此類網路資源，是當代政治、經濟、社會、文化等相關的原始性資料，是屬於「唯數位版產出」之類的資料。如為公立機構網頁提供之資料，則相當於部分圖書館慣常收集的「政府出版品」。此類網路電子資源中，學術期刊大部分並未另外發行紙本。而公、私立機構之網站資料雖可能刊印為廣告式摺頁或宣傳小冊，卻因其外形、內容未及於一般圖書館機構收存之標準，而不為圖書館持續有系統地收存典藏。因此，保存網站中的資料是此類資源保存之唯一管道。

(二)保障「開放性資源」(Open Access Resources)的永久可存取性

網路資源增加快速，消失亦快速。網際網路上每天增加約 7 百萬頁的資訊^[8]，每一個網頁的平均壽命卻僅是 44 天；1998 年建立的網頁，到 1999 年有 44% 已消失。^[9]不論是由個人或公、私立機構所建置之網站，以及網站內的資

[7] 於「中華民國期刊論文索引」中，以“Web Preservation”以及“Web Archiv*”進行查詢，查詢結果筆數為 0。而於目前國內公家網站中，亦未找到此種定期將特定網域之網站（例如所有 edu、org 等網域內的所有網站）資料備份存檔之例，故推測目前國內尚無進行網路資源備份保存之網站。

[8] “LOCKSS: About,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.

[9] Library of Congress, “Collections Policy Statement: Web Site Capture & Archiving,” (2003),

訊資源，其存在與可用性均處於隨時可能異動或消失之不穩定狀態。而網頁內容不斷更動，被取代或消失的網頁內容又可能具有研究價值。現在的網路資源之理想，是保障「開放性資源」之永久可存取性；當建置原始資料之機構異動其網頁資訊、網站連結失效或網站之服務停止時，系統化地存檔保存之數位資源備份，就成為查詢使用的重要資料來源。

(三)滿足未來可能的研究需求

Wilson (1968) 於探討書目資料製作，尤其是主題之選擇時，曾指出：「我們不能期待任何一個現在做的描述或評鑑，可以完全滿足未來的需要」，其指出的系統目標有三種層次：1. 滿足一個人達到一個目的之需求；2. 滿足一個人達到多個目的之需求；3. 滿足所有人達到各種目的之需求。達到第三種層次目標時，其能應付的面向是全面性的 (Universal Dimension)，但不易達成。^[10] 在面對網路資源，需進行收集目標之決策時，所將遭遇的兩難如同書目系統之評鑑。目前看來無什特殊的網路資源，對未來也許具有特殊研究之意義，但我們現在卻無法預知未來的需求。為了能儘可能因應未來許多人的各種不同需求，將網路資源儘可能地備份保存，成為圖書館機構蒐集購置紙本與各式實體媒體典藏之外的另一重要使命。

(四)保存當代的文化與學術資產

人類的文明不斷地演進與推陳出新，今日之常，往往成了明日之奇。現代人要了解古代音樂藝術之種種，往往憑藉的是片斷的資訊，因此只能補風捉影，無法一窺全貌，更無由得到身歷其境之感受。「現在對於希臘神話在傳承之間到底有沒有遺漏了什麼？當時的觀眾對於演出時的反應究竟怎樣？我們只能想像。對於莫札特等古典音樂大師的作品，我們也無緣知道這些作者是如何演奏自己的作品。過去英國的傑出演員 David Garrick (1717 – 1779) 深刻地影響了 18 世紀的戲劇藝術，Patrick Henry (1736 – 1799) 以他滔滔雄辯之才，成為影響美國革命的重要人物之一。但是我們也無由目睹或聆聽他們當時震撼人心的現場實境……」希望讓未來世代的人們，也能感受到現代藝術等的種種，這是 1996 年研究圖書館存取研究小組委員會 (Commission on Preservation and Access and Research Libraries

Retrieved June 1, 2006, from <http://www.loc.gov/acq/devpol/webarchive.html>.

[10] Wilson, P., "Introduction," in *Two Kinds of Power: an Essay on Bibliographical Control* (CA: Berkeley, University of California Press, 1968), pp.22-23.

Group) 討論應保存現代各種資源，包括網路上的多媒體資源時所提出的感想。

[11] 由此可知，網路資源不只包含當代社會、文化、經濟、學術等各方面之資產，也包含當代美術文藝類之資源。因此，需要保存典藏的現代資訊，除了文獻資源之外，對於現代藝術的製作過程、表演現場、網路上的各類現代藝文實況等，亦應有系統性地加以保存。

四、各國網路資源備份保存之概況及實行模式

全球網際網路的發展，提供全球資訊資源流通分享的管道，也帶給數位資訊保存極大的挑戰。網路不斷地拓展，網路資源內容多變又極易消失，因此如何擷取、保存網路數位資源，並使之長久可存取，成為數位資源備份保存計畫的重要課題。世界上一些先進國家的國家圖書館，已就如何保存其國家的當代網路文化資產，進行數位資源備份保存之先導計畫，包含：澳洲、奧地利、加拿大、捷克、丹麥、芬蘭、法國、德國、日本、立陶宛、荷蘭、紐西蘭、挪威、瑞典、英國、美國等。各國的國家網路資源備份保存計畫之網站與相關內涵，詳見本文附錄一。[12]

上述已進行國家網路資源備份保存之大部分國家，其網路資源備份保存計畫是由國家圖書館主導相關的研究與實驗，也有的國家已和出版社建立協力合作關係，拓展其備份保存之範圍。較先進的國家如英國，已在版權方面多所考量，實際透過立法，使國家的網路數位資源之備份保存，得以有法源根據；也有的國家是以寄存電子資源之實務作法支援之。近年來更有一些跨國的合作計畫，成功地將網路數位資源備份保存之相關理論付諸實現。網路數位資源之備份保存模式大致可分為以下數類：[13]

(一)特定網域全域蒐集模式 (Whole Domain Model)

特定網域全域蒐集的模式，又稱 Comprehensive Approach，是利用網路數位

[11] Commission on Preservation and Access and Research Libraries Group, "Preserving Digital Information: Report of the Task Force on Archiving of Digital Information," (May 1996) 轉引自同註 5。

[12] 1. 同註 6。

2. National Library of Australia, "PADI: Preserving Access to Digital Information," Retrieved July 10, 2006, from <http://www.nla.gov.au/padi/topics/92.html#aus>.

[13] 同註 6。

資源收集器自動地搜尋某特定網域，加以備份保存。例如瑞典國家圖書館所進行的「Kulturarw3」計畫（Cultural Heritage Cubed，網址為 <http://www.kb.se/kw3/ENG/>），將全國的網路資源均加以備份。該計畫並未對所收集的全部國內網站加以索引。因此，如欲檢索任一網站各時期的備份，必須完全正確地鍵入網址，方能進行有效查詢。Kulturarw3 並未於網路上開放查詢服務，而只限於瑞典國家圖書館館內查詢。但是對於瑞典國家圖書館之網頁，則對外開放作為範例。^[14] 另一個全域蒐集模式的典範是美國的「Internet Archive」計畫，網址為 <http://www.archive.org>，此網站採取最具雄心的全域蒐集模式，其目標是網羅保存全球的網路數位資源。

（二）選擇性蒐集模式（Selective Model）

選擇性蒐集模式是事先確定所要蒐集的網路資源範圍，訂定資源範疇之選擇政策，可能根據資源的特殊性、品質、主題、或其為某類資源群組之一部分等而定。澳洲的 PANDORA（Preserving and Accessing Networked Documentary Resources of Australia，網址為 <http://pandora.nla.gov.au/index.html>），是此種選擇性蒐集模式之典範。PANDORA 是最早開始進行網路資源備份保存的計畫之一，創立於 1996 年，由澳洲國家圖書館所主導。該蒐集計畫是透過與九所澳洲圖書館及文化保存機構合作進行的。^[15] PANDORA 訂定有極為完備之相關政策，也製備完善的相關說明與手冊，美國國會圖書館的 MINERVA 即參考其各種政策與方法，而進行網路資源備份保存之計畫。^[16]

（三）主題式蒐集模式（Thematic Model）

主題式蒐集模式是選擇性蒐集模式的一種，主要是針對特殊主題或事件之網路資源進行備份保存，例如美國國會圖書館放眼數位未來，以建立教育系列資源為職志，其由多學科背景之工作人員組成任務小組，包括：編目、法律、資訊服務、技術服務等專才，負責選擇、收集、提供服務、保存維護等工作，以服務當代與未來之世代。其 MINERVA 計畫（Mapping the INternet Electronic Resource Virtual Archive，網址為 <http://lcweb2.loc.gov/cocoon/minerva/html/minerva->

[14] National Library of Sweden, "Kulturarw3- Long Time Preservation of Electronic Documents," Retrieved July 4, 2006, from <http://www.kb.se/kw3/ENG>.

[15] National Library of Australia and Partners, "PANDORA: Australia's Web Archive," Retrieved July 4, 2006, from <http://pandora.nla.gov.au/about.html>.

[16] 同註 9。

home.html)，是以總統競選相關網站之蒐集為先導計畫，拓展蒐集主題至 2000、2002、2004 年之選舉、911 事件、冬季奧運、伊拉克戰爭、國會等，僅就該等主題進行相關網站之蒐集，至 2006 年 6 月即已蒐集 3 萬 6 千個網站。^[17] 但是此一網站數目，在 OCLC 所估計的網站數目中，僅佔 1% 而已。^[18]

(四)寄存式蒐集模式 (Deposit Model)

網站資源之寄存式蒐集模式目前尚屬實驗階段，目前尚無任何國家明確以法律規定網站所有者必須寄存網站資料。但是一些國家的出版社，或遵照國家的寄存規定、或出於自願地將數位出版品寄存至國家的網路數位資源典藏部，例如瑞典立法規定電子期刊的寄存政策，以及荷蘭的電子期刊出版社與國家的數位資源備份保存計畫達成寄存合意，二者都是成功的典範。

(五)混合式蒐集模式 (Combined Approaches)

因為沒有任何一種網路數位資源的備份保存模式是完全令人滿意的，因此有的國家如法國、丹麥等，乃採用混合式的蒐集模式。亦即既採用全域蒐集模式，又兼採選擇性或主題式的蒐集模式，以期達到最大的蒐集與資源組織之效益。此外，亦有大學研究機構是依據教學研究上的學科主題進行網路資源備份保存，例如德國海德堡大學 (Universität Heidelberg) 的「中國學研究數位資源備份保存計畫」(Digital Archive for Chinese Studies, 簡稱 DACHS)。該計畫亦收納個人針對特殊主題所蒐集的數位資源，例如由 Christan Goebel 蒐集的「臺灣民主化的基盤 2000-2001」(Taiwan's Democratic Consolidation, 2000-2001)，備份保存了超過 1 千份的各種主題之網路資源。^[19]

表二為代表性的網路資源備份保存計畫之資源保存方針、存檔頻率以及存檔項目。^[20]

[17] 同註 9。

[18] Arms, William Y., "A Report to the Library of Congress: Web Preservation Project, Interim Report," (2001), Retrieved June 1, 2006, from <http://www.loc.gov/minerva/webpresri.pdf>.

[19] Goebel, Christan, Taiwan's Democratic Consolidation, 2000 - 2001, (2003), Retrieved July 4, 2006, from <http://www.sino.uni-heidelberg.de/dachs/goebel030214.htm>. 此為 C. Goebel 為撰寫碩士論文而收集之資料，依資料之主題整理排列，包括下列之主題：Taiwanese National Identity、the International Environment Institutional Reforms、Legal Reforms、Financial Reforms、Local Factions、Corruption and Vote-buying、Party Platform Development、Parliamentary Bargaining Processes、Elite Backgrounds、Public Attitudes。

[20] 同註 5。

表二：網路資源備份保存政策比較表

機 構	保 存 方 針	存 檔 頻 率	存 檔 項 目
LC MINERVA	選擇性 selective	不定時	全部
Internet Archive	批次 bulk	每個月	HTML +images
PANDORA	選擇性 selective	不固定	全部
Kulturarw ³	批次 bulk	持續地毯式搜尋存檔 sweeps	全部

資料來源：Arms, B. & Members of the LC Web Preservation Project Team, 2001。

選擇性蒐集模式需要由館員進行選擇、索引、組織資料，使之成為專藏資料（Collection），並裨益長久保存與未來的檢索存取。但是，因為網路上的資源龐大，人工的選擇、索引、組織不可能窮其盡；故 Internet Archive 與瑞典的 Kulturarw³ 採取自動批次（Bulk）全面搜尋網路資源的方式，進行網路資源之存檔備份。自動批次全面備份保存的方式較具效率與經濟效益，可以將網路資源較完整地保存下來。^[21] 其收集資料後的編目亦是採自動化的方式，但是只依據網址、日期等進行資料組織。^[22] 惟 Internet Archive 亦與美國國會圖書館 MINERVA 合作，於美國大選時期，除一般的定期搜尋備份之外，特就競選網站之部份進行存檔保存，也於 911 事件發生後，將所備份保存之相關網站資料，提供予 MINERVA 建置主題網站。^[23]

由此可知，美國國會圖書館雖然未採用全域蒐集模式，而只進行主題模式的網路資源備份保存，但是，同樣在美國的 Internet Archive 則是採用全域模式，而且是全球網路資源為目標範圍的備份保存模式；Internet Archive 又與美國國會圖書館充分合作^[24]，因此可說是採用混合式蒐集模式的國家，只是將備份保存的職責，分散至不同的機構而已。

[21] Arms, William Y., "Coalition for Networked Information: Preserving the Open Access Web," Retrieved July 22, 2006, from <http://www.cs.cornell.edu/wya/LC-web/CNI-LC-web.ppt>.

[22] "Internet Archive: WayBack Machine- Advanced Search," Retrieved July 16, 2006, from <http://web.archive.org/collections/web/advanced.html>.

[23] "MINERVA: September 11 Web Archive- About," Retrieved July 16, 2006, from <http://memory.loc.gov/cocoon/minerva/html/sept11/sept11-about.html>.

[24] "Internet Archive: Home," Retrieved July 4, 2006, from <http://www.archive.org/index.php>.

五、網路資源備份保存與整編實務

網路資源有「開放式資源」(Open Access Resources)與「閉鎖式資源」(Close Access Resources)二種。「開放式資源」指對於資料之存取無所限制，使用者可以免費瀏覽使用並下載之網路資源。「閉鎖式資源」包括二大類，一種是原來未充分整編、保存的機構所屬資料，以大學或研究機構的學術研究成果為代表；一種則是指有存取上管控之資料庫，例如以使用者帳號密碼管控之電子期刊資源。一些已進行網路資源備份保存計畫之國家，其目標除了備份保存「開放式資源」之外，其學術研究機構亦透過「機構典藏」(Institutional Repository)之建置，保存機構的研究成果，同時促成原來處於「閉鎖」狀態的學術資源之開放分享。此外，有的國家之學術團體與電子期刊出版者結成聯盟，為保障「閉鎖式資源」之永久典藏與分享而共同合作。可知網路資源之備份保存，因數位資源之類別不同，需透過不同之實務作法，以達成保存與提供使用之目的。茲以實例說明上述各類網路資源備份保存計畫。

(一)「開放式」網路資源

「開放性資源」對於個人之瀏覽使用與機構之擷取保存，均未設限。美國國會圖書館的網路資源存檔備份保存計畫只以「開放性資源」為收集與研究之對象。蒐集保存網路資源之過程，可能包含：全自動、半自動、人工等不同的作業。從網站備份保存至提供查詢使用之相關工作，包括：訂定選擇標準與擷取政策、備份保存網路資源、編目組織擷取備份之資料、提供服務等。分述如下：

1. 訂定選擇標準與擷取政策 (Selection and Crawling Policy)

推動網路資源備份保存計畫，需先訂定選擇標準與擷取政策。茲以美國國會圖書館 MINERVA 為例，簡介選擇標準與擷取政策需考量的基本事項。MINERVA 對於各不同主題網站之備份保存，各訂立適用之選擇標準與擷取政策。以下為針對 107 屆國會網站所訂定之選擇標準與擷取政策：^[25]

Selection Criteria：

Scope: includes official Web sites of members of Congress and congressional committees. (以公家網站為範疇)

Collection Period: One-time capture on Dec. 12, 2002 (單次擷取保存之日期)

[25] “107th Congress Web Archive, About: Selection Criteria,” Retrieved July 10, 2006, from <http://memory.loc.gov/cocoon/minerva/html/107th/about-selection.html>.

Acquisition Parameters:

- Depth : The complete Web site, if possible. (講究深度，儘可能完整地擷取保存整個網站之資料)
- Breadth : Links followed out to one external level. (講究廣度，保存網站對外連結的第一層連結)
- Frequency : One-time capture (對於該網站之存檔頻率；該例為僅「單次擷取」(相對於定期重複擷取之模式))
- Robots.txt blocked sites (搜尋根目錄中嵌有 Robots.txt file 的網站)^[26]
- Web sites were crawled that used robots.txt blocking (有使用 Robots.txt file 規範搜尋引擎之索引擷取程度，而曾經被搜尋過之網站)

在選擇原則方面，可以區分為以人工選擇特定網站、以及選定特定網域而備份該網域內的所有網站資料等二類。以人工選擇所要備份保存的網站，雖然有可以作精緻選擇之優點，但是對於突發性的事件，則會措手不及，錯失備份保存之時機。此種狀況，如欲回溯備份，就只有求諸於隨時定期全面蒐集網站，且加以備份保存的網站機構了。美國發生 911 恐怖攻擊事件時，網路訊息幾乎無時無刻不在更新，因此，其 Internet Archive 對於網站之備份保存頻率次數，較平時頻繁數倍；911 恐怖攻擊事件之例，其每日對同一網站之重複備份次數，有多達 7 次者（2001 年 11 月 2 日）。MINERVA 為完整備份此類特殊事件之網路資源，將之納入其主題網站中，乃與 Internet Archive 合作，取用 Internet Archive 已存檔備份之該類網站資料，並為各個網站編製目錄。例如前述的「911 事件網站目錄」，一筆目錄乃是針對一個網址所製作，而點選該目錄中的 Archived Site，即連至由 Internet Archive 所存檔之各次存檔日期與更新情形的列表，可供點選瀏覽各次所備份保存之網站。

2. 網路資源備份保存、編目組織與服務^[27]

(1) 擷取瞬間之網站資料 (Collecting : Making a Snapshot)

利用 Mirroring Program 下載「擷取瞬間之網站資料」(Snapshot)，儲存於檔

^[26] Robots.txt file 是限定搜尋引擎可否搜尋擷取網站資訊的簡單程式檔，透過安裝該程式檔，可限定搜尋引擎只可以擷取部分網站之資訊，或容許擷取全部網站之資訊，亦即指示搜尋引擎可以索引或擷取一個網站之程度。關於 Robots.txt file 可參見“Robotstxt.ca,” (2006), Retrieved July 12, 2006, from <http://www.robotstxt.ca/index.html>.

^[27] 同註 5。

案櫃 (Archive) 中。此處所謂「擷取瞬間」是指網路資源蒐集器 (Harvester 或 Web Crawler) 搜尋而擷取下來那一片刻的資料。網站中的資料會不斷變更, 但是各網站在擷取之後的變更, 只能待下次再度搜尋蒐集 (Harvest/ Crawl) 時再加以備份保存。美國國會圖書館的 MINERVA 與澳洲的 PANDORA 等對於網站之備份保存, 都是使用免費的網站資源擷取軟體 HTTrack Mirroring Program, 進行選擇性網站資源擷取保存。^[28]HTTrack 軟體至 2006 年 6 月已發展至第 3.41-BETA-2 版, 該測試版完整的名稱為 HTTrack Website Copier, 於網路上提供免費下載。^[29]

(2) 定期擷取存檔 (Collecting : Periodic Snapshots)

依照設定的存檔頻率, 如: 每月、每週、每天下載網站資料, 也就是定期擷取存檔, 累積不同次的「擷取瞬間之網站資料」, 將之一一儲存於檔案櫃中。

(3) 由電腦就所擷取之資料進行分析 (Analysis by Computer)

利用電腦程式分析各次所下載的網站資料, 包括與上一次存檔的版本作比較, 判斷是否有更新, 以及自動擷取相當於 Metadata 之資料, 例如: 網站名稱、日期、網站所屬機構 (網站之作者) 等。如果是主題式蒐集模式之網站, 對於該等資料則會經過人工確認, 再將之存入典藏機構之系統中。

(4) 使用查詢者之分析選擇 (Analysis by Patron)

不論是「全域蒐集模式」或「選擇性蒐集模式」之網路資源備份, 都可以設計為讓使用者自己評估或挑選所要使用的版本。例如 Internet Archive 是直接列載各次存檔之不同版本, 提供點選瀏覽。美國國會圖書館的 MINERVA 計畫則提供較多的選擇, 包括在「整合式圖書館系統」(Integrated Library System, 簡稱 ILS) 中檢索所需瀏覽使用的網站, 提供可以瀏覽的、已存檔之網站網址清單、題名清單、主題清單等^[30], 提供使用者點選後連至已備份保存的舊網站資料。

(5) 資料編輯、編目與儲存典藏 (Cataloging, Storage and Preservation)

完整的網路資源備份保存流程包括存檔資料的登錄 (Accession Control)、編目等, 才能提升查詢的效益。網路資源存檔備份與提供服務之完整流程包括: 網

[28] 1. 同註 18。

2. Phillips, Margaret, "PANDORA, Australia's Web Archive, and the Digital Archiving System that Supports It," (2003), Retrieved June 1, 2006, from <http://www.nla.gov.au/nla/staffpaper/2003/mphillips1.html>. 亦見 *DigiCULT.info* 6 (Dec. 2003), p. 24.

[29] "HTTrack Website Copier," Retrieved July 10, 2006, from <http://www.httrack.com/>.

[30] 同註 18。

路資料擷取、登錄、電腦分析、編目、儲存以至提供使用。其中登錄與編目是提供使用者查詢利用的關鍵性工作。另外，經備份保存後的網路資源，究竟要提供「原始未經編輯的下載版本」(Snapshot Files)，或是提供「編輯後的讀者存取用的版本」(Access Version)，端視機構的服務政策而定。而服務機構決定提供何種版本，則是依據該機構之營運目的與使用該資源的社群而定。例如 Internet Archive 營運之目的是為保存全人類知識，所提供的是屬於原始未經編輯的下載版本，以供各行各業進行查詢。美國國會圖書館 MINERVA 計畫之營運目的，基於保存「系列性的教育資源」之考量，對於所擷取備份保存之網路資源，會進行較多包含智識判斷的資訊組織，提供的是屬於「編輯後的讀者存取用的版本」。至於「編輯後的讀者存取用的版本」在編輯方面要做到何種程度，也是依機構之政策而定。可以只進行些微的編輯，也可以進行全面的編輯，以符合特定的主題或使用習慣；可以全自動地進行編輯，也可透過人工進行編輯。表三是 Internet Archive、PANDORA、MINERVA 三個計畫讀者版本編輯方式之比較。^[31]

表三：Internet Archive、PANDORA、MINERVA計畫讀者版本編輯方式之比較

網站名稱	比較項目	編輯方式	編輯與否
Internet Archive		電腦	電腦自動作業後，不再編輯
PANDORA		研究人員	人工編輯
MINERVA		研究人員	人工編輯

資料來源：Arms, B.&Members of the Le web Preservation Project Team, 2001。

在編目工作方面，下載後的網路資源之編目可分為二種，一種為全自動，另一種為半自動處理方式。半自動處理方式是指透過自動收集方式所擷取之資料，以及經過自動編目之資料，需再經過人工處理。

搜尋備份保存政策與編目政策，影響成本甚鉅。Internet Archive 的自動批次收集講究廣度，以全自動的方式存檔及提供利用。美國國會圖書館 MINERVA 講究深度，以人工進行所下載之資料的整理編目。成本方面，美國國會圖書館 MINERVA 高於 Internet Archive 大約 100 倍。^[32] 由此看來，基於作業效率的

[31] 同註 5。

[32] 同註 18。

考量，進行網路資源存檔備份的網站機構，最好具備以電腦程式分析所下載的網站資料之功能，並能自動進行基本的編輯，例如先自動編製目錄，使人工的編輯能減至最少，以節約成本。

(6) 使用者介面設計 (Service : Search Interface)

網路資源經備份保存於檔案櫃後，可以透過查詢系統之建置與查詢介面設計提供利用。使用者查詢介面所提供的查詢方式，可以依資料性質而作不同的設計。以下為常見的查詢功能設計：

① 提供所備份保存的網站之網址清單 (List of Sites by URL, 例如, MINERVA: Alphabetical Listing of September 11 Web Archive Contents)。^[33]

② 提供網址結合日期之查詢 (Access by URL + Date, 例如 Internet Archive: WayBack Machine—Advanced Search)。

③ 提供以自動作業方式所完成之索引 (Automatic Index, e.g., Web Search Engines, 例如 MINERVA 中的「第 107 屆國會網站備份」(107th Congress Web Archive) 提供索引檢索)。^[34]

④ 提供特定議題 (Topic) 之主題網站清單 (Subject Access), 例如：PANDORA。

⑤ 提供特定議題之網站名稱清單 (List of Sites by Title), 例如：

A. MINERVA 中的第 107 屆國會之網站 (107th Congress Web Archive), 其提供相當於網站名稱的國會議員人名供點選, 網址為 <http://memory.loc.gov/cocoon/minerva/html/107th/browse-person.html> (檢索日期：2006 年 7 月 9 日), 並提供圖書館「整合式圖書館系統」的目錄檢索。其製作目錄時索引之檢索點包括議員所來自的州名、人名、委員會名稱等, 網址為 <http://memory.loc.gov/cocoon/minerva/html/107th/index.html> (檢索日期：2006 年 7 月 9 日)。

B. PANDORA 主題清單點選進入後, 第二層即提供網站名稱, 例如研討會名稱或期刊名稱。

⑥ 提供 MARC 或 Dublin Core 形式之目錄, 例如 MINERVA 為每一個所存

[33] “MINERVA: Alphabetical Listing of September 11 Web Archive Contents,” Retrieved July 16, 2006, from http://www.loc.gov/minerva/collect/sept11/full_list.html.

[34] “107th Congress Web Archive,” Retrieved July 9, 2006, from <http://memory.loc.gov/cocoon/minerva/html/107th/index.html>.

檔備份之網站，均製作詳細的目錄，參見：

A. MINERVA 中，國會議員 Tom Allen 網站之目錄，網址為 <http://lcweb2.loc.gov/cocoon/minerva/0192/default.html>（檢索日期：2006 年 7 月 10 日）：提供有關該議員之個人相關描述性資料之外，並提供連結至其個人網站之資料。由於個人網站中包含許多子網頁，因此是屬於有關個人之「完整典藏為單位的目錄」（Collection Level Catalog）。

B. MINERVA 中，American Anglican Council 所建置的 911 事件網站之目錄，網址為 <http://lcweb4.loc.gov/911/catalog/0121.html>（檢索日期：2006 年 7 月 9 日）：提供不同方式的瀏覽查詢，包括：網站架設者、網站所屬機構或團體的性質、網站所屬國家等之瀏覽清單。而有關 911 事件的一個網站之目錄，則提供多次重複備份之資料。

⑦ 提供單一主題網站或關於某主題之群組網站之目錄（Catalog Record for Individual Site or Group of Sites，例如 Archives Hub: Revolutionising Access to the Archives of UK Universities and Colleges— A National Gateway to Descriptions of Archives in UK Universities and Colleges”）。^[35]

⑧ 提供圖書館目錄查詢（Access Through Library Catalog），例如 MINERVA 中所保存的網站，可以在國會圖書館目錄中查得。

由上述可知，開放式網路資源之備份保存，是透過自動或半自動的方式蒐尋備份保存，而在提供利用方面，如欲提供有良好組織的備份資料予讀者利用，則有賴於圖書館員的加值，可說是圖書館業務之新領域。MINERVA 對於網路資源之編目，是利用 OCLC 的 CORC System，製作 MARC Catalog Records，再將之載入圖書館的「整合式圖書館系統」，編目的過程頗類似期刊之編目。

網路資源之保存計畫可能會就相同的網址，定期重複地進行存檔備份。對於所擷取存檔備份之網站，應給與基本的編目資料以茲識別，並提供檢索。各網路資源備份保存計畫之網站，對於所存檔資源的編目，可能各訂有一定的目錄格式（Metadata Schema）。以 MINERVA 為例，其所使用的「保存目的之 Metadata」（Preservation Metadata）之格式為 Metadata Object Description Schema

[35] “Archives Hub: Revolutionising Access to the Archives of UK Universities and Colleges- A National Gateway to Descriptions of Archives in UK Universities and Colleges,” Retrieved July 10, 2006, from <http://www.archiveshub.ac.uk/>.

(簡稱 MODS)，目前已發展至 3.0 版，該 MODS 3.0 為 XML 格式之架構，提供需要建檔的書目元素，幫助建置來源資料之目錄。^[36]MOD 對於網路資源編目之基本資料，規範必須具有識別 (Identification) 以下二者之作用：^[37]

- A. 識別網站：識別網站有賴於 URL 或 URN (例如 Handle 或 PURL)，但是如前所述，許多網站可能變更其 URL，而需要人工進行新、舊 URL 之間的關聯。
- B. 識別來源與版本：為了明示來源與版本，需要著錄網站的識別碼 (Web Site Identifier)、網站相關訊息，例如擷取存檔之日期與時間 (Collection Information, e.g., Date, Time, etc.)、網站變更之歷史 (History of Changes) 等。

但是，網路資源編目往往有需要人工輔助解決的問題，例如自動分析製作題名時，常有不易取擇之難，主要是因為許多網站的 HTML <title> 並未完善製作。另外，識別資料 (Identifiers) 經常有多重識別資料 (Multiple Identifiers) 與 URL 更改的問題。因為很難確定對使用者最好的編目方式為何，因此，此類網路數位資源之最佳編目方式並不易確定。不過美國國會圖書館對於網路資源之編目，抱持肯定的態度；認為以完整的一個典藏為單位的目錄 (Collection Level Catalog) 十分有用，值得製作。^[38]有鑑於此，網路資源之製作單位，例如各種網站之建置者，實需要對嵌入 Metadata 於網頁中的問題，更確實嚴謹地執行。

(二)「閉鎖式」網路資源

「閉鎖式資源」(Close Access Resources) 之備份保存，或因擁有資源之機構尚未系統性地建置資料庫保存之，或因原來的網站建置者如出版社，其對於資料庫均設有帳號密碼之管控，因此並非一般的搜尋所得以擷取備份。

但是，除了「開放性資源」有必要進行備份保存，以保障永久的可存取性之外，對於「閉鎖式資源」亦有必要爭取其納入國家級、或是永久性機構之營運保存系統中，方符合「災難防範」(Disaster Plan) 與確保穩定且長久可存取之原則。

國外先進國家對於網路上的「開放性資源」，大多由國家級機構主導備份保

[36] "MINERVA: Metadata Component," Retrieved July 16, 2006, from <http://memory.loc.gov/cocoon/minerva/html/minerva-metadata.html>.

[37] 同註 5。

[38] 同註 5。

存，但是「閉鎖性資源」因關係到內容建置方面需要統籌管理機制，例如出版社所建置之電子期刊全文系統、各圖書館與博物館等典藏機構之數位典藏計畫、學校與研究機構之研究成果，大多是由個別機構自行建立存檔備份的保存機制。美國的 LOCKSS Alliance（LOCKSS 一詞來自 Lots of Copies Keep Stuff Safe）、英國 JISC 主導的 UK LOCKSS Pilot Programme、澳洲國家圖書館之將網路上的「閉鎖性資源」納入國家的數位資源保存計畫中^[39]等，都是代表性例子。此外，亦有大學基於保存學校教師與研究人員之研究教學成果，建置學校的機構典藏系統以保存之，例如美國麻省理工學院（簡稱 MIT）建置的機構典藏 DSpace。此等個別機構的備份保存機制，將可成為國家級、或是其他永久性網路資源備份保存機構之基礎。以下茲就「閉鎖式資源」的「機構典藏」與「電子期刊連盟」等二大類，分別舉例說明之，以作為我國建立更多此類保存計畫之參考。

1. 機構典藏（Institutional Repository）

大學是學術研究的重鎮，也是研究與教學成果產出的主要園地，舉凡研究上的論文、教學講義素材、研究計畫成果報告等，與電子期刊一般，均為重要的學術資產。此類資料如彙集成資料庫，推廣使用而服務相關的學術社群，將對人類社會之進步具有重大貢獻。以下為該類計畫之例：

（1）美國麻省理工學院建置之DSpace收錄該校博碩士論文，至2006年7月止，已收錄11,000篇論文，並提供教師自行將個人著作之電子版加入該系統中。^[40]

（2）瑞典Lund University Library所建置之Directory of Open Access Journals（簡稱DOAJ）。將網路上的全文電子期刊建置為資料庫網站，該資料庫收錄多語文之電子期刊，至2006年7月止，已收錄2,292種期刊，共包含101,489篇論文，其中有653種期刊已可以檢索期刊內的單篇文章。^[41]該資料庫網站訂定有嚴格的電子期刊遴選標準^[42]，成為極具權威性與知名度之開放存取電子期刊網站。

（3）臺灣大學圖書館主導建置之「中文電子學位論文服務」（Chinese

[39] “Policy and Practice Statement : Collecting Policy,” Retrieved July 17, 2006, from http://pandora.nla.gov.au/policy_practice.html.

[40] MIT Libraries, “Dspace,” Retrieved July 2, 2006, from <https://dspace.mit.edu/index.jsp>.

[41] Lund University Libraries, “Directory of Open Access Journals (DOAJ),” (Sweden) Retrieved July 2, 2006, from <http://www.doaj.org/>.

[42] Lund University Libraries, “Directory of Open Access Journals (DOAJ): About,” (Sweden), Retrieved July 2, 2006, from <http://www.doaj.org/articles/about#background>.

Electronic Theses & Dissertations Service，簡稱CETD），以該校博碩士論文全面數位化為目標，執行「臺大博碩士論文資料庫建置計畫」。此一計畫徵求本校博碩士校友，將學位論文以非專屬授權的方式授權予學校建置資料庫，以充實學校之學術資料館藏，並推廣至全球學術機構，希冀提升臺灣大學的學術影響力。^[43]自2004年6月至2006年7月止，在7,604位畢業生中，已有6,252位完成論文授權，授權率高達82.22%。^[44]臺灣大學更以此為基礎，主導「中文電子學位論文服務」學位論文資料庫之建置，收錄臺灣大學、淡江大學、中山醫學大學、中興大學、臺北科技大學、元智大學、高雄醫學大學、中國醫藥大學、臺北藝術大學、長榮大學、清華大學等校的博碩士論文，目前該資料庫總計已收錄 21,791 篇博碩士論文之書目資料，全文篇數已達 13,194 篇之多。^[45]除了就博碩士論文進行數位典藏外，臺灣大學亦由圖書館規劃建置類似MIT的DSpace之臺灣大學機構典藏系統，以典藏保存全校教師與研究者之研究成果。該系統目前尚屬實驗階段，未來推出，將可作為臺灣其他機構建置同類機構典藏之參考。

但是不論是機構典藏或電子期刊典藏網站之建置，同樣都可能遭遇個人或機構基於版權之考量而拒絕之阻礙。〈臺灣機構典藏發展芻議〉一文提出「三權漸進」之主張，將有關著作之權利分為作者自我典藏之權（Self Archiving Right）、全文索引之權（Full Text Indexing Right）、以及取用之權（Access Right）等三者，該文指出，在此一模式之下，作者提供著作予機構典藏，作者除了自己的著作獲得永久典藏、提高自己著作之可見度等好處外，有關個人權益之部分亦可得到保障。^[46]此一概念，當有助於機構典藏之實現。

2. 電子期刊連盟

電子期刊是學術研究成果最重要的傳播媒體之一，也是學術界高度需要的研究資源。但是因為其關係到出版社之利益，其傳播利用上的藩籬亦較高。有鑑於

[43] 項潔，〈臺大博碩士論文資料庫建置計畫〉。上網日期：2006.12.5。網址：<http://www.lib.ntu.edu.tw/etds/home.htm>。

[44] 〈臺大博碩士論文資料庫建置計畫〉。上網日期：2006.12.5。網址：<http://www.lib.ntu.edu.tw/etds/home.htm>。

[45] 〈中文電子學位論文服務：瀏覽說明〉。上網日期：2006.12.5。網址：<http://www.etds.com.tw/ec/browse.aspx> 2006/12/5。

[46] 項潔、洪筱盈，〈臺灣機構典藏發展芻議〉，《教育資料與圖書館學》，43：2（2005.12），頁 173-189。

電子期刊資源是重要的學者研究結果之結晶，負載人類過去與未來之智慧，而且暢通使用管道有利於學術研究之進步提升，先進國家或由國家機構、或由學術團體，陸續提出相關的期刊資源保存計畫。以下為國外電子期刊資源保存之相關計畫：

(1) EVA-the Acquisition and Archiving of Electronic Network Publications

為芬蘭 Helsinki University Library 於 1997 年在其國家教育部的支援下執行之計畫。該計畫由圖書館、出版社、專門性機構等共同合作，主要目標是要開發供蒐集與著錄網路出版品之工具。目前的具體成果包括：Muisti-project for Digitization and Network Delivery of the Finnish National Collection^[47]，以及 Elektra-Finnish Scientific Publications in Electronic Form^[48]，其中 Elektra 為電子期刊之保存計畫，收錄 40 種科學性期刊全文計 13,000 篇（大部分為 pdf 檔），以及少部分學術論文之全文。該資料庫僅提供付費訂購之機構使用全文，大部分的芬蘭大學圖書館均有訂購此資料庫。

(2) SHERPA-Securing a Hybrid Environment for Research Preservation and Access

英國的諾丁罕大學（University of Nottingham）在 Joint Information Systems Committee（簡稱 JISC）與 Consortium（簡稱 CURL）之支援下，建置「保障研究資源存用成長環境」網站（Securing a Hybrid Environment for Research Preservation and Access，簡稱 SHERPA）。^[49] 瑞典的 DOAJ 與英國的 SHERPA 二者合作整合其電子期刊資源，建置 Directory of Open Access Repositories（簡稱 OpenDOAR）。^[50] 而 CURL 另也支援 JISC-CURL LOCKSS Pilot Programme，此計畫利用美國史丹佛大學圖書館所開發的免費電子資源搜尋備份軟體，是透過與出版社合作的電子期刊資源保存計畫。^[51]

(3) LOCKSS—Lots of Copies Keep Stuff Safe

[47] “Muisti-project for Digitization and Network Delivery of the Finnish National Collection,” Retrieved July 15, 2006, from <http://www.lib.helsinki.fi/memory/etusivue.html>.

[48] “Elektra - Finnish Scientific Publications in Electronic Form,” Retrieved July 15, 2006, from <http://www.lib.helsinki.fi/elektra/english.html>.

[49] “University of Nottingham, UK. SHERPA: Home,” Retrieved July 2, 2006, from <http://www.sherpa.ac.uk/>.

[50] “The Directory of Open Access Repositories – OpenDOAR,” Retrieved July 2, 2006, from <http://www.opendoar.org/>.

[51] “JISC. UK LOCKSS Pilot Programme,” Retrieved July 2, 2006, from <http://www.jisc.ac.uk/index>.

① LOCKSS 的歷史由來^[52]

LOCKSS 為全國性的電子期刊備份保存計畫。網際網路的資訊包含不少「唯數位版產出」的電子期刊，因此是圖書館所必要努力蒐集的資料之一。此類資料必須即時備份保存，因為其產出之後，很可能一週或數天之後即自網路上消失。史丹佛大學圖書館因此主導開發可以幫助圖書館即時蒐集備份保存此類資料的軟體。該軟體名稱來自「多備份保障安全」的英文 Lots of Copies Keep Stuff Safe (簡稱 LOCKSS)，旨在協助圖書館在無太大的經費負擔之下，進行其訂購的網路資源之蒐集備份保存，建立自己的網路資源資料庫，以保障數位資源的永久可即性；因此以烏龜為其 logo，取其長壽之意。LOCKSS 可以安裝於個人電腦中，個人電腦即成為數位資源保存之工具，可以在最低費用之下，於電子期刊一刊出後即將之備份保存，成為未來可提供利用的資源，而無虞網路上的資源消失。如此，圖書館即可保障其讀者永遠可以存取該電子期刊之備份保存檔。LOCKSS 設計上充分考量資料備份的完整性，以及系統的安全性，保障使用該軟體的圖書館與出版社等之系統安全。

1999 至 2000 年間為 LOCKSS 軟體測試期，於 2000 至 2002 年之間，已成功地在世界上 50 所圖書館安裝運作，當時尚需要些微的人力介入。2002 至 2004 年間，Emory 大學圖書館、印地安那大學圖書館、以及紐約公共圖書館等的館員，就館藏發展、館藏管理、以及存取使用等已提出無數問題，供 LOCKSS 進行改善；LOCKSS 系統終得以在 2004 年 4 月正式推出。目前全世界已有超過 80 家圖書館、50 家出版社，共同結合成 LOCKSS 聯盟 (LOCKSS Alliance)，共同努力確保需要訂購、以及非營利性質之電子期刊的永久可存取性。LOCKSS 研究開發與維運經費的主要來源是 Andrew W. Mellon Foundation 與 National Science Foundation，另外也從英國 Joint Information Systems Committee (簡稱 JISC)、企業界 Sun Microsystems 與 HP Labs、以及學界如：Intel Research Berkeley、Stanford University Libraries and Academic Information Resources、Stanford Computer Science Department、Harvard Computer Science Department 等得到部分經費支援。LOCKSS 之研發是由 Michael A. Keller 計畫主持，其研究開發團隊包括 6 位美國當地、一

cfm?name=programme_lockss. 此為為期二年之計畫，執行期間為 2006 年 3 月至 2008 年 2 月。

[52] “About LOCKSS,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.

位來自英國 Digital Curation Center (Glasgow) 的工程師，以及包括來自柏克萊大學、哈佛大學、Hewlett Packard Laboratories, (加州) 的 6 位研究員。^[53]

② LOCKSS 的設計理念^[54]

LOCKSS 的設計對於圖書館、出版社、以及研究者均有特別的考量。對於圖書館是希望能幫助其簡易且經濟地進行出版、保存、並備份屬於其圖書館的電子館藏，使圖書館能夠擁有，而不需於未來再進行訂購當下的該批資源，如此既可實現圖書館傳統的保存任務，又可保障圖書館對於現有的該批資源的永久可存取性；對於出版者，LOCKSS 考量的是協助其方便且經濟地將內容提供予圖書館進行備份保存，並保障其經營模式 (Business Models) 與出版平台可能招致的風險降至最低，另方面亦可保障使用者長久存取其資料，滿足圖書館對於所訂購的電子資源永續存取之要求。而 LOCKSS 對於研究者與電子期刊之讀者的考量，則是使讀者在無感覺 LOCKSS 存在的情境下，得以利用現成可用的搜尋引擎，無礙地查詢被圖書館備份保存之數位資源，而且能夠在出版者的原始 URL 中無礙地存取新出版的、以及備份版的電子資源。基於上述為圖書館、出版社、使用者之各種考量，LOCKSS 技術設計上的基本理念包括以下數點：

- A. 經濟性：數位資源保存上最大的困難之一是經費上的需求；許多典藏機構都為維護保存經費不足所困擾，因此需使必要的經費需求盡量降低，使保存上的經營既可成長，又可長久運作，不致於因為營運費過高而中途停頓。
- B. 建立使用者與資助研發者之信心：如果數位資源被備份保存之後，即有如進入黑箱一般，不為一般使用者所頻頻利用，則不易評估其正確性與可用性，而無法建立讀者與經費資助者之信心。因此數位資源的備份保存系統，應成為使用者於網路檢索瀏覽上的常設部分，提供隨時查詢及存取，才能建立使用者與資助研發者之信心。
- C. 以 Pull 模式取代 Push 模式，以提升效益性：為了降低圖書館與出版社使用該系統之成本，應使需要進行網路數位資源備份保存的典藏機構可以利用備份保存的工具 (Preservation Appliances)，進入出版者的網站搜尋 (Crawl)，進而擷取出資料，此即所謂的「Pull」模式。「Pull」

[53] “About LOCKSS: Team,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.

[54] 同註 51。

模式所擷取備份保存的是出版者原已設計完好的網路「呈現版」(Presentation Form)，典藏機構再將之提供予讀者時，仍然一如出版者出刊時之樣貌，讀者不會感到不習慣或不便，此為較傳統「Push」模式更為經濟有效的方法。傳統的「Push」模式是出版者將資料庫存置於第三者(Third-party，例如 Mirror Site 或代理商)的典藏庫中。此種模式對於出版者而言，等於允許第三者訂定新的流通目的(Re-purpose)與重新發行(Re-publish)，風險更大。而對於訂購該電子資源的圖書館典藏機構而言，因為必須再進行網頁的製作，以重現出版者的檢索介面，其所需花費的成本也較高。

- D. 讓使用者端成為一個 Proxy，使其查詢及存取無礙：讓使用者端成為一個 Proxy，使用者查詢時所在的位址仍為原始的 URL，因此其原設定的「書籤」(Bookmarks)與檢索指令等，於查詢備份檔時仍舊有效。此對於使用者而言，無疑是更為經濟有效的。
- E. 讓圖書館之間建立合作機制，使其電子期刊資源保存之成本降低：不同的圖書館就同樣的電子期刊進行備份保存，則任何一份期刊文章的檔案故障時，圖書館之間可以互相支援，省卻圖書館需要隨時檢測系統內檔案資源的有效性之功夫。雖然各機構整體上所需的硬體設備總數較多，但是因為各機構所需的硬體為屬於消費者層級者，而非業界層級者，而且也不需花費太多行政成本，故可降低整體上的成本。再者，圖書館典藏機構以 LOCKSS 系統模式進行所訂購或免費訂閱的電子期刊之備份後，並無需另立預算項目，故不會有被刪除預算而中斷營運之風險。

③ LOCKSS 聯盟之營運政策與使用社群

目前 LOCKSS 軟體之使用者機構，已結合成 LOCKSS 聯盟 (LOCKSS Alliance)，共同為備份保存電子期刊與網路數位資源而努力。其利用 LOCKSS 軟體而備份保存之數位資源包括：博碩士論文、美國聯邦政府與政府之檔案文件、各種特殊資料、數位影像與電子書、Blogs、網站等。^[55] 該聯盟採會員制，各會員除將機構本身所製作產出之數位資源提供予該聯盟會員存取之外，亦有繳交會

[55] "LOCKSS: Publishers and Titles - Electronic Journals Released for Collection and Preservation," (July 12, 2006), Retrieved July 15, 2006, from http://www.lockss.org/lockss/Publishers_and_Titles.

費之義務。而該聯盟亦隨時發佈最新上線提供存取之數位資源的訊息，以及新加入的會員名單與會員簡介。該聯盟將所擁有之數位資源，區分為會員提供者以及「開放性資源」二類，由會員提供之資源，唯有繳交會員費之會員機構，才能進入查詢存取資料。此外會員尚可得到下載額外的數位資源所需的軟體、可以被及早告知 LOCKSS 的新技術與政策、取得正式公開前之軟體、參加各種工作坊、洽詢 LOCKSS 專門技術人員之機會等。另一方面，圖書館典藏機構在數位資源創造上的努力成果，亦因透過此聯盟之公開，而能得到更好的推廣傳播之效。^[56] LOCKSS Alliance 之營運極為成功，最近甫得到 OCLC 之加入。^[57]

與 LOCKSS Alliance 相關的另一組織為 CLOCKSS (Controlled Lots of Copies Keep Stuff Safe)，此為非營利之組織，其成立宗旨是在結合圖書館與出版社為夥伴關係，希望相關政策乃是由社群成員共同決定^[58]，保障有效、完整的數位典藏，以滿足未來對學術性數位資源的存取需求；其成員包括 American Medical Association, American Chemical Society, American Physiological Society, Blackwell Publishing, Elsevier, Institute of Physics, John Wiley & Sons, Inc. Nature Publishing Group, Oxford University Press, SAGE Publications, Springer, Taylor and Francis 等出版者，以及 Indiana University, New York Public Library, OCLC, Rice University, Stanford University, University of Edinburgh, University of Virginia 等的圖書館機構。^[59] 美國國會圖書館肯定該組織之貢獻與重要性，因此，核准將補助 70 萬美元。^[60]

由於美國史丹佛大學圖書館所開發的此一 LOCKSS 軟體，已發展至相當成熟之境地，因此，英國的資訊系統聯盟 (Joint Information Systems Committee，簡

[56] "LOCKSS: LOCKSS Alliance," Retrieved July 15, 2006, from http://www.lockss.org/lockss/LOCKSS_Alliance.

[57] "LOCKSS: News Archive (6/23/06)," Retrieved July 15, 2006, from <http://www.lockss.org/lockss/Home#BePress>.

[58] "CLOCKSS: Home - About CLOCKSS," Retrieved July 15, 2006, from <http://www.lockss.org/clockss/Home>.

[59] "CLOCKSS: Home - Participating Members of CLOCKSS," Retrieved July 15, 2006, from <http://www.lockss.org/clockss/Home>.

[60] "CLOCKSS: Library of Congress Announces Digital Preservation Award to Stanford University," (June 7, 2006), Retrieved July 15, 2006, from http://www.lockss.org/clockss/Library_of_Congress_Press_Release.

[61] "UK LOCKSS Pilot Programme," Retrieved July 15, 2006, from http://www.jisc.ac.uk/index.cfm?name=programme_lockss. 亦見 http://www.jisc.ac.uk/fundingopportunities/funding_calls/2005/12/

稱 JISC) 亦採用此軟體，主導英國電子期刊之備份保存計畫。該計畫設有技術支援部門，對於參與圖書館，提供技術支援，並與出版社洽商，徵集其電子全文。出版社同意之情況，將會簽署書面以及可機讀之同意書；並將其系統設定為允許 LOCKSS 軟體進入資料庫系統中搜尋最新出版之資料，進行備份保存。^[61]

由此可知，電子期刊資源之備份保存，有賴跨機構以及圖書館與出版社之合作。先進國家嘗試電子期刊之整編計畫雖然始於 1990 年代末期，但是備份開放性電子期刊 (Open Access Journals) 之理念，則是在 2002 年於 Lund Copenhagen 舉行的斯堪的納維亞學術傳播研討會中首次被討論。^[62] 而在短短三、四年之間，先進國家圖書館對開放式與閉鎖式電子期刊之備份典藏計畫，已具相當成效，值得借鏡。

六、臺灣網站為國外網路資源備份保存網站保存之情形

國內雖然尚未有國家機構執行定期備份保存網路資源之計畫，但是，因為 Internet Archive 是以全球為蒐集範圍，因此臺灣目前已不存在之歷史性網頁，可以由 Internet Archive 之網站查詢而得。透過在 Internet Archive 的 WayBack Machine 中試查臺灣的各類機構之網頁，可以進行臺灣各類機構網站的各種統計研究，並可瞭解臺灣網站為國外網路資源備份保存網站保存之情形。

(一) Internet Archive 簡介

Internet Archive 創立於 1996 年，位於 San Francisco，其創立目的是為提供全面存取人類知識之管道。該網站之資料來源有由該網站自行搜尋保存者，也有來自其他典藏機構寄存者；自 1999 年起接受 Alexa Internet 以及其他機構及個人網站直接給予資料，而成長為更完備的網路資源保存網站。Internet Archive 保存之網路資源種類包括：文字資料、影音資料、動畫、軟體、網站等。2003 年，Internet Archive 將所備份保存之資料全數寄存一份於埃及新建的 Library of Alexandria，相當於建置一個 Mirror Site。當 San Francisco 的 WayBack Machine 當機時，可改於 Library of Alexandria (網址為 <http://archive.bibalex.org>) 的 WayBack

funding_circular7_05.aspx。

[62] 同註 41。

[63] 在埃及 Library of Alexandria 的 WayBack Machine 之網址為 <http://www.bibalex.org/english/initiatives/internetarchive/web.htm> (Retrieved July 16, 2006)。

[64] “Internet Archive: About AI,” Retrieved June 1, 2006, from <http://www.archive.org/about/about.php>.

Machine 查詢。^[63]但是，Library of Alexandria 的 Mirror Site 所提供之資料少了約兩年。Internet Archive 的合作機構包括：維吉尼亞大學（Library of Virginia）、多倫多大學（University of Toronto）、印地安那大學（Indiana University）、北卡羅萊納州的典藏機構（North Carolina State Archives），以及其他多個學術機構與紀念館。^[64]該網站從 2000 年至今大約五年半期間，存檔的網頁數量從 10 億頁成長為 550 億頁，大約成長了 55 倍^[65]；網路資源成長之速由此可見。

Internet Archive 的蒐集備份保存政策是，限定網域為 edu、gov、org 等為蒐集範圍，此外亦選擇性地收集一些 com 之網站。另外該網站亦透過 “Archive-It 1.5”，收費性地接受個人網站備份申請。從提供網址給 Internet Archive 到可以在 WayBack Machine 查得所被保存之資料，約需 6 個月的等待時間。^[66]

(二)臺灣網站為Internet Archive存檔備份之情形

為了解臺灣網站被 Internet Archive 備份保存之情形，作者於 Internet Archive 的 WayBack Machine（網址為 <http://www.archive.org/index.php>，檢索日期：2006 年 6 月 11 日）中，試查臺灣 edu、gov、org、com 等各類網域之網站被存檔情形。作者試圖以 54 個臺灣網站（網域為 gov 18 個、org 8 個、com 1 個以及 edu 27 個）進行統計，於 WayBack Machine 中逐一鍵入所欲查詢舊網頁資料的網站網址，所得結果清單中，列有各年度、各次搜尋存檔之日期。該網站之系統如果偵測到網頁較上一次存檔有所更動，則以星號「*」加以標識。作者將各次查詢之網址、Internet Archive 對該網址備份之日期起訖、備份起訖日期之間的備份存檔次數、該網址內容更動（網站之自我更新）次數等，一一鍵入 Excel 工作表中，然後統計各網站每個月平均被存檔次數，以及各網站每個月自我更新之平均次數，如附錄二所示。

經作者統計，所試查的 54 個網站，從 1996 年至 2005 年，分別自不同時期開始進行備份保存。此 54 個網站，在過去 9 年之間，共被存檔 8,220 次，平均每個月網站每個月被存檔次數是 1.9 次，而各網站內容的更動頻率則是每個月平

[63] 此為由 WayBack Machine 查得的 2000 年之 Internet Archive 首頁所提供的統計資料（網址為 <http://web.archive.org/web/20010118201300/http://www.archive.org/>）與 2006 年 6 月該網站的首頁（網址為 <http://www.archive.org/index.php>，Retrieved June 1, 2006）所提供的統計資料的比較。

[66] 同註 24。

均 0.7 次。此外，不同類型網域之網站被存檔備份的頻率，以教育類為最高，為每個月平均 2.2 次。其餘類型之網域則相差不多，每個月大約平均 1.4 至 1.5 次，詳如表四所示。

表四：臺灣網站為Internet Archive備份保存之樣例統計

網站類型	月數	存檔 次數 Results	平均存檔 次數 次數/月	更新 次數 Updates	平均更新 頻率 次數/月
gov	1231	1890	1.5	630	0.5
org	536	759	1.4	338	0.6
com	56	84	1.5	26	0.5
edu	2551	5487	2.2	1939	0.8
gov_org_com_edu	4374	8220	1.9	2933	0.7

資料來源：本文作者研究整理。

WayBack Machine 查詢結果所呈現的清單中，當年度之資料均未提供點選瀏覽，顯示該網站對於其蒐集範圍之網域，雖然定期搜尋（Crawl）備份保存，但是從存檔後到可提供檢索舊網頁之間，大約有一年的等待期。

Internet Archive 的蒐集政策是定期自動搜尋備份保存。作者為了解其對於含有各種圖形之網頁的存檔情形，乃於國立臺灣大學網站的備份保存清單中，試著點選瀏覽 1997 年、1999 年及 2005 年的過期網頁，結果發現，大部分的網頁均被有效地存檔，網頁中的連結也大致有效，但是部分圖形則無法呈現，顯見其未被存檔。其中 2005 年的圖書館網頁中的同一式項目符號圖形，有的可呈現，有的卻不可呈現，此應是因為該類圖形之存檔，所需時間超過其設定之時間範圍，因「逾時」（Time Out）而被中途放棄。

不同的網路資源備份保存網站機構之政策不同，所能存檔之資料內容亦有差異。一般而言，影響自動收集保存之因素包括：

1. 以資料庫的形式提供之資料；
2. 以通用性不佳的軟體繪製的圖形；
3. 無效的連結；
4. 自動批次收集的情形，有的設定不保存聲音檔與動畫檔；
5. 圖形過多或圖檔過大，以至於超過存檔備份機構所設定的自動蒐尋下載之

「逾時」(Time Out) 時間，致使存檔不完整。

由此可知，網頁的設計如果使用通用性低的圖形製作軟體，包含多種媒體資料，則被完整存檔之可能性隨之降低。而網站之內由帳號密碼管控之資料庫，則大多不可能被存檔。

爲了未來網頁與網站資料能被充分備份保存，各機構於網頁設計、網站架構方面，必須使之符合「網站可即」(Accessibility) 之原則。目前已有許多軟體公司，開發網頁設計的檢核工具，以幫助製作符合「網站可即」原則之網頁與網站。易於使用的有：Watchfire® Bobby™ 所提供的網頁檢核工具，名稱爲 WebXact。^[67] 於 WebXact 鍵入網址執行檢測後，WebXact 將列出檢測的結果；檢測結果分爲 General、Quality、Accessibility 及 Privacy 等部分。“General” 項目就網頁之內容作一摘錄，包括：檔案大小、最近更新日期、Metadata、網頁內包含的表格與圖形數目等。此處的 Metadata 是指從網頁的 HTML 原始檔案所能獲取的 Metadata 資訊。該 Metadata 資訊之完整與否，影響到網路資源存檔備份後的自動編目成效。“Quality” 項目所提供的資訊主要包括：內容的錯誤（例如無效的連結）、檢索與導航功能、網頁的效能（例如於使用者連入該網頁時提供警示語）、網頁使用不同瀏覽器時的有效性等。

如前所述，網站被存檔備份時，原來網頁連結的有效與否完全照常被存檔；原來無效的連結，備份保存後，該失效的連結就無法修正，而使網站的備份保存不完整。因此此部分檢核之結果，對於改善網頁以利完整備份，有所幫助。至於“Accessibility” 項目，則就檢測所發現的問題，將問題分爲“Priority 1”（必須修改）、“Priority 2”（宜修改）、“Priority 3”（可修改）等三類。經測試多個國、內外圖書館機構的網站，發現其實網站並不容易完全通過三個層級之檢測，但是，通過“Priority 1”之檢測則是必要的。有關各層次問題，以及其對網頁可即性之影響，可參照 W3C 的「Web Content Accessibility Guidelines 1.0」。^[68]

除了 Watchfire® Bobby™ 提供網頁檢核工具之外，Etre's Accessibility Check

[67] WebXact 提供檢核網頁之網址爲 <http://webxact.watchfire.com/> (Retrieved July 16, 2006)。

[68] W3C, *Web Content Accessibility Guidelines 1.0*. (1999), Retrieved July 16, 2006 from <http://www.w3.org/TR/WAI-WEBCONTENT/>.

[69] Etre's Accessibility Check 在網路上提供網頁檢核工具之網址：<http://www.etre.com/tools/accessibilitycheck/> (Retrieved July 16, 2006)。

[70] “The History of A-Prompt,” Retrieved July 16, 2006, from <http://aprompt.snow.utoronto.ca/history.html>.

與 A-Prompt 同樣都是根據 W3C 的「Web Content Accessibility Guidelines 1.0」所開發。Etre's Accessibility Check 也如 WebXact 一般，在網路上直接提供網頁檢核工具。^[69]A-Prompt 是由美國威斯康辛大學的 TRACE Center 與加拿大多倫多大學 Adaptive Technology Resource Centre（簡稱 ATRC）共同開發完成^[70]，於網路上提供免費下載。^[71]此外，於網路世界中最大的開放性軟體資源網站 SourceForge.net 中，以“Accessibility Software”為關鍵詞進行檢索，可以查得各種功用略有不同的網頁可即性檢核軟體。^[72]網站架設機構宜善加利用此類檢核工具，以使網路資源備份保存達到完整備份的最佳效果。

七、結論與建議

網路資源備份保存的目的，在於幫助保存流傳於網路世界中的文化與學術資產，對於連結研究學者與歷史性資料，具有重要功能。但是網路資源之保存成效，有賴於資料建置端與資料備份保存端的雙軌努力，作者根據研究本議題之心得，歸納結論與建議如下：

(一)結論

1. 國家的網路資源備份保存計畫需要各界協力

網路資源之存檔備份，需要有可長可久的實務上之維護與營運。Google 雖採全域蒐尋保存之政策，以建置廣泛規模資料庫為目標，但因其宗旨不在永久典藏，每月淘汰數位資料多達 200 萬頁^[73]，故不足以作為網路資源永久典藏之憑藉。對於網路資源保存典藏最具保障的，乃是由負長久保存之責的國家或長久性機構擔負。先進國家之學者曾呼籲應由國家擔當此責任^[74]，實際上，現在已有不少國家確實是由國家級之機構主導與負國家網路數位資源保存之責。但是網路資源之存檔備份，涉及的層面甚廣，除了需要有長久性的經費支援之外，亦需要

[71] 軟體 A-Prompt : Web Accessibility Verifier 目前已發展至 A-Prompt Version 1.0.6，Retrieved July 16, 2006, from <http://aprompt.snow.utoronto.ca/overview.html>。該軟體可以自以下網址下載：<http://aprompt.snow.utoronto.ca/download.html>（Retrieved July 16, 2006）。

[72] 於 SourceForge.net 網站中，以“accessibility Software”進行檢索之結果。Retrieved July 16, 2006 from http://sourceforge.net/search/?words=Software&type_of_search=soft&words=accessibility+Software&Search=Search。

[73] 同註 21。

[74] National Research Council, *A Digital Strategy for the Library of Congress* (National Academy Press, 2000), Retrieved June 1, 2006, from <http://www.nap.edu/books/0309071445/html/>。

資訊科技、資訊選擇與組織等方面人才之襄助。國家網路數位資源之備份保存既需要多方面專業的結合，由國家級機構主導並負典藏維護之責，同時結合技術面較具新科技能力的學術界，以及資訊選擇與組織人力較優越的圖書館機構，應可使此計畫之推展，更具完備性與效益。

2. 網頁設計與網頁基本資訊應考量網站備份之簡易完整性

進行網路資源備份保存之機構，其對網路資源之保存工作，或基於現實上的人力與經費條件之考量，或基於其所設定的服務目的，均設有存檔備份上的政策，例如完整地全部存檔備份，或只下載儲存特定格式之檔案等。而基於下載存檔效率之考量，亦可能有逾時即停止下載之設定。因此一個機構如欲網站被完整存檔，應以網站易於被完整存檔備份為設計上的考量。由本文對於國立臺灣大學歷年來網頁備份之查詢實例可知，減少使用圖形或使用通用性高的製圖軟體，有助於網站之被完整備份。此外，為便於網站被下載後的自動編目，網站應加入網頁與網站的 Metadata 資訊。而網站設計是否符合標準理想，建議利用檢核工具，掃描檢核所設計的網頁，確認是否符合網頁設計之原則，至少通過基本驗證（Priority 1 的項目）之檢核方正式上載至網路，開放查詢使用。

3. 參與機構需要適度重整人力資源組織

圖書資訊組織之工作需要相當的人力。以美國國會圖書館的 MINERVA 為例，其投入的人力包括美國國會圖書館 7 人、康乃爾大學（Cornell University）1 人、Internet Archive 計畫 3 人，此 11 人團隊為全時工作的專家團隊。^[75]即使是採全自動蒐集備份的 Internet Archive，亦有 7 位全時工作人員參與。至於每天有 600 萬以上檢索人次的 Google，其雖採全域備份保存之政策，人力介入之需要相對較少，但人力亦達 80 餘人之多，其中 10 餘人為具有電腦科學博士學位之專業。^[76]可知妥善處理網路資源之存檔備份，以及選擇、組織編目已備份保存之網路資源，均需要相當專業人力之投入，因此，參與機構宜適度重整人力資源組織，以期工作之質優效佳。

4. 面對著作權法上的新課題，思考解決之道

網路資源之備份保存，是符合教育學界與國家整體利益之事業，值得肯定支持。但是此一事業之目的是在於保障開放性資源，甚至是閉鎖性資源之長久的、

[75] 同註 5。

[76] 同註 21。

開放性的使用管道。著作權相關法規之更新腳步，遠不及資訊建設理念之發展，以至於有些國家在存檔備份網路資源，並加以整理編目之後，仍遲疑於將之全面開放，例如瑞典 Kulturarw³ 並未於網路上開放查詢服務，而只限於瑞典國家圖書館館內查詢。美國的學者亦曾撰文陳述著作權相關法規不明確，造成數位時代之困境。^[77] 我國開始規劃網路資源存檔備份保存計畫之初，實亦需要與著作權法之管理機構進行充分討論，為我國備份保存之網路資源的傳播利用，開創低障礙之利用環境，使此類計畫之努力，得以發揮最高效用。

(二)建議

1. 臺灣應自行建立網路資源備份保存計畫

網路資訊環境的發達，造成網路數位資源量增，臺灣不僅在資訊基礎建設方面媲美先進國家，在網路資訊內容的發展方面，也不落人後。過去幾年國家投注經費於數位典藏，成就了國家級與地方級典藏機構大量數位化其珍貴的文物收藏。此外，公、私立機構與個人亦多利用網路科技，傳播各種訊息與個人文藝創作。今日的網路資源，尤其是「唯數位版產出」資料，都將成為未來研究現今社會、政治、經濟、文化、教育等的重要第一手資料。由本文之介紹可知，不少先進國家，或以國家圖書館為主導，或輔之以大學機構的大力協助，或由國家訂定相關寄存法規，期使國家的數位資源保存達到理想境地。臺灣至今為止，尚無此種國家級的網路資源保存計畫，甚為可惜。因此，基於保存當代文化資產之考量，實應及時規劃及進行國家的網路資源備份保存計畫。

2. 結合專家學者研商國家網路資源備份存檔的範疇、相關政策與分工

本文所介紹的「全域」蒐集模式、「選擇性」蒐集模式、「主題式」蒐集模式、「寄存式」蒐集模式、「混合式」蒐集模式等，優異性互見，也各有執行上的問題。一般而言，對國家具有重要性的資料，如：國家各公部門之網站資料、已有紙本但收存其同步發行的數位資源有益保存及推廣使用之資料、對學術具有權威性與研究價值之資料、反應當代文化之代表性資料等，都有必要加以保存。但基於完整性與經濟效率，宜考慮兼採批次「全域蒐集保存」之模式與「選擇性

[77] 1. National Research Council, *The Digital Dilemma, Intellectual Property in the Information Age* (National Academy Press, 2000), Retrieved July 22, 2006, from http://www.nap.edu/html/digital_dilemma/.

2. 同註 21。

及主題性蒐集保存」之模式，亦即採用混合式之模式，除了批次選擇收集特定類別的網站（例如 edu、gov、org 等網域之網站），亦個別選擇已知的重要網站，作深入的蒐集（因為重要人物的個人網站網域可能是 com），並選擇有關教育上的系列主題與重要時事之網站，加以備份保存及整理為主題群組網站，裨益研究利用。^[78]但是大型計畫之前期規劃，將關係到永續經營之成效，故宜結合專家學者，就國家網路資源備份存檔的範疇、對不同類型網站定期備份之間隔時間、分工上的相關政策等進行討論，以保障規劃完善可行。

3. 圖書館機構與出版界合作，建立夥伴關係

出版界的紙本與數位出版品，尤其是有專業評審制（Peer Review）之出版品，是資訊資源的重要來源，尤其是電子期刊、學術論著的網路典藏資料庫（Scholarly Articles Repository）等，是研究成果的豐富典藏，因此在保存國家數位資源計畫中，不宜偏廢此一部分。但是，如果是營利性質的電子出版品或是其資料庫，則網路資源備份保存後的開放性服務，是與私人出版界的營利方針相抵觸的。如何在兩全的考量下，使出版界與圖書館攜手合作，讓國家級典藏機構，為現代數位資源之備份保存增添一份保障，值得圖書館界與出版界共同思考，研商可行之策。美國的 LOCKSS Alliance 建立會員制模式，並以 CLOCKSS 結合圖書館與出版界，共同監督政策訂定，以保障符合社群之共同利益，獲得許多知名大學、出版社、以及各種公、私立機構加入支持，此一成功範例，值得參考。

4. 建立國際合作關係，爭取 Internet Archive 過去備份保存之臺灣網路資源

國外的網路資源備份保存，開始於 1995 年，亦即視窗版網際網路興起後不久。如本文之介紹，Internet Archive 對臺灣網路資源之備份乃是始於 1996 年。Internet Archive 起步較早，確實對網路資源備份保存貢獻匪淺。而 Internet Archive 起始階段，得到埃及亞歷山卓圖書館寄存資料而建立初步規模。此一模式足堪效法。如果能夠與之建立合作關係，得到其過去備份保存之臺灣網路資源，則可更完整地建立國家的網路資源典藏，可說遺珠復得，為後代留下可貴的現代文化資產。

5. 圖書資訊學教育應將網路資源備份保存之技術與相關議題列入課程之中

選擇性及主題性蒐集保存之模式，需要具有圖書館學選擇資訊之訓練的館員支援。此外，網路資源備份保存後亦需自動、半自動、或人工製備各網站備份之

[78] 同註 18。

目錄，以便提供查詢檢索；惟有結合網路資訊科技與資訊組織之專業，才可能圓滿達成任務。但是，因為技術方面已有成熟的免費公用軟體可資參考利用，而資訊選擇及組織編目則尚大力仰賴人力智識之判斷，此為新的網路時代中，圖書館員之重要技能。學者 William Y. Arms 視之為圖書館學與資訊科學之間的新學域 (Librarianship in a New Domain)，而呼籲圖書館界應參與擔負國家網路資源備份保存與組織整理之任務。^[79] 為使我國的圖書資訊與網路資源整編之人才，足敷我國網路科技與資訊資源增長中對專才之需要，圖書資訊學教育應與時俱進，一方面增加人才之培育，另方面則應將網路資源備份保存之技術與相關議題列入課程之中，使國家資源保存政策之推動，得有充分的人力為其柱石。

6. 加入會員制的數位資源聯盟，以有效增益國內的可用數位資源

本文所介紹的 LOCKSS Alliance，結合圖書館界與出版界，以會員制的方式，提供繳交會員費之會員，共享該聯盟所建置之系統內的各種資料庫數位資源，會員也可以取得數位資源以及其他各種資訊。此種會員制的數位資源聯盟模式，既保障出版者之權益，又可整合、保存數位資源，不但值得國內效法，而且也值得國內機構試行加入，以有效增益國內的可用數位資源。但是，國內各機構先結成聯盟，以團體機構之組織加入，或可得到更佳效益。在加入此類聯盟的同時，因會員身份而提供予該類聯盟的數位資源，也將獲得推廣之效。

7. 圖書館應嘗試了解備份所訂購電子期刊之可行性

本文中提到，LOCKSS 之設計理念之一，是讓圖書館之間建立合作機制，使圖書館電子期刊資源保存之成本降低。其實際作法是不同的圖書館就同樣的電子期刊進行備份保存，則任何一份期刊文章的檔案故障時，圖書館之間可以互相支援。此外，圖書館典藏機構進行所訂購或免費訂閱的電子期刊之備份後，未來即不虞期刊刪訂之問題。國內學術機構於訂閱電子期刊後，是否能夠將之全面備份，以及國內要以怎樣的分工合作機制進行備份，值得進行深入了解與探討。

總之，網路數位資源不論對現代或是未來的學者都具有相當的參考研究價值，而因為網路資源有其不穩定性，其備份保存宜及早實施，以免造成無可彌補的遺憾。由本文之探討可知，網路資源之備份保存是否能夠完善進行，所需要的努力是多方面的：負責蒐集備份保存網路資源的典藏機構，必須持恆地定時向外徵集，亦即定時搜尋所設定為範圍的網域，將網路資料備份保存，同時以自動

^[79] 同註 21。

或部分配合人工的半自動方式，進行資源的編目組織。而爲了對「開放式資源」與「閉鎖式資源」都同樣充分備份保存，圖書館界需要努力與學界及出版界溝通合作。再者，所有產出數位資源的機構，對於自己機構之網頁與網站設計，以及提供於網路上的資源都需重新審視修正，不論在圖形利用、連結的有效性、「保存目的之 Metadata」（Preservation Metadata）之完整性等方面，都使之符合資源可即性原則；如此多管齊下的配合與努力，才可能使國內未來的網路資源備份保存計畫順利執行，而保障現今與未來世代的民衆與研究學者，不論對於「開放式資源」或「閉鎖式資源」，都能夠存取使用，不虞資源在不經意之間倏爾消失。

致 謝

本文爲作者在國立臺灣大學圖書資訊學研究所修習博士班課程「數位典藏專題研討」時之研究，承蒙授課教師陳雪華教授與臺灣大學圖書館項潔館長之指導與鼓勵甚多。此外，不具名審稿委員們亦提供寶貴的建議，謹在此一併敬致最深謝忱；另，作者自 2000 年起參與國科會之「國家典藏數位化計畫」、「數位典藏國家型科技計畫」、以及文建會「國家文化資料庫計畫」所補助之計畫的執行，對於數位資源之永續備份保存十分關心，僅以此文，獻給奉獻心力於國家文化資產數位典藏的機構與工作夥伴。

（收到日期：95.7.23；修訂接受日期：95.12.14）

參考文獻

- 「中文電子學位論文服務：瀏覽說明」。上網日期：2006.12.5。網址：<http://www.etsd.com.tw/ec/browse.aspx>。
- 「臺大博碩士論文資料庫建置計畫」。上網日期：2006.12.5。網址：<http://www.lib.ntu.edu.tw/etsd/home.htm>。
- 項潔。「臺大博碩士論文資料庫建置計畫」。上網日期：2006.12.5。網址：<http://www.lib.ntu.edu.tw/etsd/home.htm>。
- 項潔、洪筱盈。〈臺灣機構典藏發展芻議〉，《教育資料與圖書館學》，43：2（2005.12），頁173-189。
- “107th Congress Web Archive, About: Selection Criteria,” Retrieved July 10, 2006, from <http://memory.loc.gov/cocoon/minerva/html/107th/about-selection.html>.

- “107th Congress Web Archive,” Retrieved July 9, 2006, from <http://memory.loc.gov/cocoon/minerva/html/107th/index.html>.
- “About LOCKSS,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.
- “About LOCKSS: Team,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.
- “Archives Hub: Revolutionising Access to the Archives of UK Universities and Colleges- A national Gateway to Descriptions of Archives in UK Universities and Colleges,” Retrieved July 10, 2006, from <http://www.archiveshub.ac.uk>.
- Arms, B. and Members of the LC Web Preservation Project Team, “LC's Web Preservation Project: a Pilot for Selecting, Collecting, Cataloging, and Accessing Archived Websites: PPT, at Mumford Room, Madison Building,” (2001), Retrieved July 16, 2006 from <http://www.loc.gov/minerva/presentations/minervafeb2001.ppt>.
- Arms, William Y., “A Report to the Library of Congress: Web Preservation Project, Interim Report,” (2001), Retrieved June 1, 2006, from <http://www.loc.gov/minerva/webpresi.pdf>.
- Arms, William Y., “Coalition for Networked Information: Preserving the Open Access Web,” Retrieved July 22, 2006, from <http://www.cs.cornell.edu/wya/LC-web/CNI-LC-web.ppt>.
- “CLOCKSS: Home- About CLOCKSS,” Retrieved July 15, 2006, from <http://www.lockss.org/clockss/Home>.
- “CLOCKSS: Home - Participating Members of CLOCKSS,” Retrieved July 15, 2006, from <http://www.lockss.org/clockss/Home>.
- “CLOCKSS: Library of Congress Announces Digital Preservation Award to Stanford University,” (June 7, 2006), Retrieved July 15, 2006, from http://www.lockss.org/clockss/Library_of_Congress_Press_Release.
- Commission on Preservation and Access and Research Libraries Group, “Preserving Digital Information: Report of the Task Force on Archiving of Digital Information,” (May 1996).
- “DigiCULT.info” 6 (Dec. 2003).
- “Elektra - Finnish Scientific Publications in Electronic Form,” Retrieved July 15, 2006, from <http://www.lib.helsinki.fi/elektra/english.html>.
- Goebel, Christan, Taiwan's Democratic Consolidation, 2000 – 2001 (2003), Retrieved July 4, 2006, from <http://www.sino.uni-heidelberg.de/dachs/goebel030214.htm>.
- “HTTrack Website Copier,” Retrieved July 10, 2006, from <http://www.httrack.com>.
- “Internet Archive : WayBack Machine-Advanced Search,” Retrieved July 16, 2006, from <http://>

- web.archive.org/collections/web/advanced.html.
- “Internet Archive: About AI,” Retrieved June 1, 2006, from <http://www.archive.org/about/about.php>.
- “Internet Archive: Home,” Retrieved July 4, 2006, from <http://www.archive.org/index.php>.
- “Internet Archive: Universal Access to Human Knowledge,” Retrieved June 1, 2006, from <http://web.archive.org/index.php>.
- “Internet Resources,” Retrieved June 1, 2006, from <http://library.ust.hk/res/beyond>.
- “JISC. UK LOCKSS Pilot Programme,” Retrieved July 2, 2006, from http://www.jisc.ac.uk/index.cfm?name=programme_lockss.
- “Library of Alexandria: WayBack Machine,” Retrieved July 16, 2006, from <http://www.bibalex.org/english/initiatives/internetarchive/web.htm>.
- Library of Congress, “About MINERVA,” Retrieved June 1, 2006, from <http://lcweb2.loc.gov/cocoon/minerva/html/minerva-about.html>.
- Library of Congress, “Collections Policy Statement: Web Site Capture & Archiving,” (2003), Retrieved June 1, 2006, from <http://www.loc.gov/acq/devpol/webarchive.html>.
- “LOCKSS: LOCKSS Alliance,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/LOCKSS_Alliance.
- “LOCKSS: News Archive,” (July 12, 2006), Retrieved July 15, 2006, from <http://www.lockss.org/lockss/Home#BePress>.
- “LOCKSS: Publishers and Titles- Electronic Journals Released for Collection and Preservation,” (July 12, 2006), Retrieved July 15, 2006, from http://www.lockss.org/lockss/Publishers_and_Titles.
- “LOCKSS: About,” Retrieved July 15, 2006, from http://www.lockss.org/lockss/About_LOCKSS.
- Lund University Libraries, “Directory of Open Access Journals (DOAJ),” (Sweden) Retrieved July 2, 2006, from <http://www.doaj.org>.
- Lund University Libraries, “Directory of Open Access Journals (DOAJ): About,” (Sweden), Retrieved July 2, 2006, from <http://www.doaj.org/articles/about#background>.
- “MINERVA: Metadata Component,” Retrieved July 16, 2006, from <http://memory.loc.gov/cocoon/minerva/html/minerva-metadata.html>.
- “MINERVA: Alphabetical Listing of September 11 Web Archive Contents,” Retrieved July 16, 2006, from http://www.loc.gov/minerva/collect/sept11/full_list.html.
- “MINERVA: September 11 Web Archive- About,” Retrieved July 16, 2006, from <http://memory>.

- loc.gov/cocoon/minerva/html/sept11/sept11-about.html.
- MIT Libraries, "Dspace," Retrieved July 2, 2006, from <https://dspace.mit.edu/index.jsp>.
- "Muisti-project for Digitization and Network Delivery of the Finnish National Collection," Retrieved July 15, 2006, from <http://www.lib.helsinki.fi/memory/etusivue.html>.
- National Library of Australia and Partners, "PANDORA: Australia's Web Archive," Retrieved July 4, 2006, from <http://pandora.nla.gov.au/about.html>.
- National Library of Australia, "PADI: Preserving Access to Digital Information," Retrieved July 10, 2006, from <http://www.nla.gov.au/padi/topics/92.html#aus>.
- National Library of Australia, "Web archiving," Retrieved June 1, 2006, from <http://www.nla.gov.au/padi/topics/92.html>.
- National Library of Sweden, "Kulturarw3- Long time preservation of electronic documents," Retrieved July 4, 2006, from <http://www.kb.se/kw3/ENG>.
- National Research Council, A Digital Strategy for the Library of Congress (National Academy Press, 2000), Retrieved June 1, 2006, from <http://www.nap.edu/books/0309071445/html>.
- National Research Council, The Digital Dilemma, Intellectual Property in the Information Age (National Academy Press, 2000), Retrieved July 22, 2006, from http://www.nap.edu/html/digital_dilemma.
- "NEDLIB: FAQ," Retrieved July 22, 2006, from <http://nedlib.kb.nl/index.html?related.html>.
- Phillips, Margaret, "PANDORA, Australia's Web Archive, and the Digital Archiving System that Supports it," (2003), Retrieved June 1, 2006, from <http://www.nla.gov.au/nla/staffpaper/2003/mphillips1.html>.
- "Policy and Practice Statement: Collecting Policy," Retrieved July 17, 2006, from http://pandora.nla.gov.au/policy_practice.html.
- "Robotstxt.ca," (2006), Retrieved July 12, 2006, from <http://www.robotstxt.ca/index.html>.
- "The Directory of Open Access Repositories – OpenDOAR," . Retrieved July 2, 2006, from <http://www.opendoar.org>.
- "The History of A-Prompt," Retrieved July 16, 2006, from <http://aprompt.snow.utoronto.ca/history.html>.
- "UK LOCKSS Pilot Programme," Retrieved July 15, 2006, from http://www.jisc.ac.uk/index.cfm?name=programme_lockss.
- "UK LOCKSS Pilot Programme," Retrieved July 15, 2006, from http://www.jisc.ac.uk/fundingopportunities/funding_calls/2005/12/funding_circular7_05.aspx.

University of Nottingham, “UK. SHERPA: Home,” Retrieved July 2, 2006, from <http://www.sherpa.ac.uk>.

W3C, Web Content Accessibility Guidelines 1.0. (1999), Retrieved July 16, 2006, from <http://www.w3.org/TR/WAI-WEBCONTENT>.

Wilson, P., “Introduction,” in *Two Kinds of Power: an Essay on Bibliographical Control* (CA: Berkeley, University of California Press, 1968).

附錄一 各國網路資源備份保存計畫

國別 / 數位資源保存計畫名稱與網址	概 要
澳洲 (Australia) PANDORA archive http://pandora.nla.gov.au/index.html	1. 由國家圖書館主導 2. 始於1996年 3. 選擇對國家有意義、對研究有用者 4. 與出版社有協定，部分資料內容需經許可才可存取
奧地利 (Austria) AOLA (the Austrian Online Archive) http://www.ifs.tuwien.ac.at/~aola	1. 1999年開始備份保存網路資源 2. 2000年7月法律規定寄存非線上的數位出版品 3. 由奧地利國家圖書館與維也納科技大學 (Vienna University of Technology) 的軟體資訊學系 (Department of Software Technology) 共同合作 4. 蒐集保存全奧地利網域 (.at) 內的資料 5. 2001年5月使用NEDLIB harvester，但因為不穩定，當年6月即改採用瑞典的Kulturarw ³ 計畫之Combine harvester，兼採全域與選擇性的備份保存，效果良好 6. 2004年因為經費不足暫停，2005年再重新啟動
加拿大 (Canada) LAC (Libraries and Archives Canada) http://www.collectionscanada.ca/index-e.html	1. 1994-1995以「電子出版品先導計畫」(Electronic Publications Pilot Project, 簡稱EPPP) 調查線上出版品之相關課題 2. 建立Electronic Collection Web site, 由各界自願提供資料；目前收集的資料包括出版業界、非營利機構、個人、聯邦政府以及省級政府機構之數位出版品 3. 2004年5月公佈開放查詢與自由存取大約總備份的80%之資料；目前資料總量與個別數位資源典藏名稱詳見< http://www.collectionscanada.ca/collection/index-e.html > (accessed 2006/7/22) 4. National Archival Development Program (NADP), 自2006年1月起提供數位化經費供申請
捷克 (Czech Republic) WebArchiv - Archive of the Czech http://en.webarchiv.cz/	1. 2000年在國家文化部 (Ministry of Culture) 之支援下開始進行相關技術之研究。 2. 由國家圖書館主導，由Moravian Library提供技術支援；Institute of Computer Science of Masaryk University 為合作機構。 3. 兼採全國網域以及特選主題與事件之網路資源的蒐集備份保存。 4. 以備份軟體工具Heritrix進行蒐集備份。 5. 目前總資料量為2 TB，約2千6百萬頁資料。 6. 只開放出版者同意再公開之資源供查詢與存取。

國別 / 數位資源保存計畫名稱與網址	概 要
丹麥 (Denmark) Netarchive.dk http://netarchive.dk/index-en.php	1. 2005年通過數位資源寄存之相關立法。 2. 由國家的The Royal Library與The State and University Library共同合作進行國家網路數位資源之備份保存。 3. 採全域 (.dk) 蒐集備份保存之政策。 4. 基於資料安全之考量，合作的兩個機構均保存備份。 5. 備份資料並未公開使用，必須取得Danish Data Protection Agency許可才可進入備份資料庫查詢存取。
芬蘭 (Finland) EVA archival system--the acquisition and archiving of electronic network publications http://www.lib.helsinki.fi/eva/english.html	由芬蘭國家教育部 (The Finnish Ministry of Education) 主導。
法國 (France) BnF http://www.bnf.fr/default.htm	1. 2001年6月開始試行蒐集備份。 2. 由國家圖書館與Institut National de l'Audiovisuel (簡稱INA，負責電視與收音機廣播節目之保存) 共同合作，進行網路資料之備份保存；無法自網路擷取保存者，則由出版者直提供寄存。 3. 兼採全域保存 (一年數次) 與選擇性蒐集備份保存 (約針對全部網站的10%進行較頻繁的備份保存)。 4. 與the French National Institute for Research in Computer Science and Automatic Control (簡稱INRIA) 進行合作。 5. 利用Metadata擷取軟體工具DeepArc擷取原網頁之Metadata資訊。 6. 所備份保存之資源尚未公開提供使用，部分可在法國國家圖書館館內使用。
德國 (Germany) Deposit.DBB.DE archive http://deposit.ddb.de/index_e.htm	1. 由國家圖書館主導。 2. 目前資料庫中包括由出版者提供之數位資料 (始於2001年9月)、博士與博士後論文 (始於1998年7月)、德國佔領時期的法律文獻 (始於2000年9月) 等。 3. 使用HTTrack harvester於2000-2001年進行網路資源蒐集備份保存，其結論是，對德國而言，以法律強制寄存比依賴網路資料搜集擷取器的“pull”方式有效。 4. 涉及版權與收費性資料僅限德國國家圖書館內使用。
日本 (Japan) WARP (Web Archiving Project) http://warp.ndl.go.jp	1. 由國家圖書館 (日本國會圖書館) 主導。 2. 2000年立法規定電子出版品如光碟之寄存，2002年日本國會圖書館提案應立法規定網路線上出版品之寄存。 3. 選擇性備份保存。 4. 蒐集備份一千餘種網路上免費的電子期刊。 5. 提供公開使用。
立陶宛 (Lithuania) Archive of Electronic Resources http://www.inforum.cz/inforum2003/prispevky/Jodelis_Remigijus.pdf	1. 由國家圖書館主導。 2. 始於2002年10月，2002年11月進行第二次測試性蒐集，採集了60種已註冊之期刊。 3. 與出版者洽商，至2003年已能蒐集備份立陶宛主要的出版社之電子出版品。 4. 現在是每半年搜尋備份一次，並即時解決版權問題，且即時擷取Dublin Core格式之Metadata資訊。

國別 / 數位資源保存計畫名稱與網址	概 要
荷蘭 (The Netherlands) The archiving system for electronic publications: e-Depot http://www.kb.nl/dnp/e-depot/dm/dm-en.html	1. 由國家圖書館主導。 2. 始於2003年，為世界上第一個規模完整的電子期刊之備份保存計畫。 3. 與出版者如Springer, Blackwell Publishing and Taylor&Francis. 合作，使e-Depot得以備份保存70%以上STM-journals (Science, Technology, Medicine) 期刊，並持續爭取出版者加入中。 4. 在「歐洲寄存圖書館網」(Networked European Deposit Library, 簡稱NEDLIB)*中為領銜者。
紐西蘭 (New Zealand) New Zealand digital archive http://www.natlib.govt.nz/en/digital/index.html http://www.natlib.govt.nz/bin/media/pr?item=1085885702	1. 由國家圖書館主導。 2. 概念來自於美國的RLG與OCLC。 3. 首次測試性的蒐集備份保存為1999年選舉網站的收集 4. 自2002年開始進行相關研究。 5. 2003年立法將寄存資料範圍拓展至包括線上與非線上資源。 6. 2004年徹底重審國家圖書館的數位圖書館之發展方向，正式宣佈將擔負起備份保存網路資源之責任。 7. 兼採國內全域(.nz)與選擇性/主題性網站蒐集備份保存之政策。
挪威 (Norway) Paradigma Project http://www.nb.no/paradigma/eng_index.html	1. 由國家圖書館主導。 2. 為三年計畫：2001年8月至2004年12月。 3. 採全域(.no)蒐集備份保存政策。 4. 2002年12月至2003年1月搜尋備份時，保存了310萬個URL之網站資料。2003年8月再次搜尋備份時，保存了410萬個URL之網站資料。約半年期間，挪威國內增加約100萬個網站。 5. 蒐集備份範圍涵蓋'.com'，'.org' and '.net'等。 6. 針對六十餘種網路電子報之網站，每日進行備份。 7. 計劃採用IFLA的「書目記錄必備功能」(Functional Requirements for Bibliographic Records, 簡稱FRBR)之模式，進行網路資源之編目。 8. 計劃採用Nordic Web Archive's (簡稱NWA) Access Tool建置提供使用者查詢資料之系統。
瑞典 (Sweden) Kulturarw³ (Cultural Heritage Cubed) http://www.kb.se/kw3/ENG	參見本文四之(一)、(五)。
英國 (United Kingdom) Domain.uk	1. 2001年由大英圖書館主導為期6個月的實驗性計畫Domain.uk。 2. 針對所選擇之網站，一一爭取網站所有者之同意。 3. 該計畫成功經驗，是現今大英圖書館網路資源備份保存計畫之基礎，使大英圖書館之該計畫，成為UK Web Archiving Consortium (UKWAC)中極重要之部分。 4. 加入由法國國家圖書館主導的International Internet Preservation Consortium (IIPC)。 5. 2004年國家對寄存資料之立法延伸涵蓋範圍，包括電子書、電子期刊、網站等，使英國之網路資源擷取備份保存能有完備的法律根據。

國別 / 數位資源保存計畫名稱與網址	概 要
美國 (United States) 1. Minerva: Mapping the Internet the Electronic Resources Virtual Archive 2. Internet Archive 3. LOCKSS Alliance: 'Lots of Copies Keep Stuff Safe' Alliance	1. Minerva 由國家圖書館主導。 2. Internet Archive 由法人組織 (.org) 主導。 3. LOCKSS Alliance 由法人組織 (.org) 主導，為包括電子期刊、數位典藏資源等之保存計畫。

資料來源：1. National Library of Australia, "Web Archiving," Retrieved June 1, 2006, from <http://www.nla.gov.au/padi/topics/92.html>.

2. National Library of Australia, "PADI: Preserving Access to Digital Information," Retrieved July 10, 2006, from <http://www.nla.gov.au/padi/topics/92.html#aus>.

說明：* "NEDLIB: FAQ," Retrieved July 22, 2006, from <http://nedlib.kb.nl/index.html?related.html>. 該計畫為期 3 年；始於 1998 年 1 月，止於 2000 年 12 月。

附錄二 臺灣網站為 Internet Archive 備份保存情形參考範例

網站類型	網站名稱	存檔起迄	月數	存檔次數 Results	平均 存檔次數 (次數/月)	更新 次數 Updates	平均 更新頻率 (次數/月)
gov	故宮博物院 http://www.npm.gov.tw/	Dec. 1997-Nov. 2004	84	180	2.1	33	0.4
gov	臺灣博物館 http://www.ntm.gov.tw	April 2004-Mar. 2005	12	24	2.0	14	1.2
gov	國家臺灣文學館 http://www.nmtl.gov.tw/	Feb. 2005-Mar 2005	2	2	1.0	2	1.0
gov	國史館 http://www.dmh.gov.tw	Feb. 1998-Mar. 26, 2005	84	98	1.2	42	0.5
gov	國立科學工藝博物館 http://kaohsiungwalking.kcg.gov.tw/	Feb 2002-March 2005	36	39	1.1	5	0.1
gov	總統府網站 http://www.president.gov.tw/	Oct. 2000-March 2005	54	104	1.9	82	1.5
gov	行政院 http://www.ey.gov.tw/	May 1998-March 2005	83	132	1.6	27	0.3
gov	立法院 http://www.ly.gov.tw/	Jun 1998-March 2005	82	109	1.3	15	0.2
gov	司法院 www.judicial.gov.tw	March 1997-March 2005	97	154	1.6	32	0.3
gov	考試院 http://www.exam.gov.tw/	Feb. 1998-April 2005	85	97	1.1	35	0.4
gov	監察院 http://www.cy.gov.tw/	Dec. 1998-March 2005	76	56	0.7	15	0.2
gov	行政院國家科學委員會 http://web.nsc.gov.tw/	Aug. 2004-April 2005	9	55	6.1	11	1.2
gov	行政院文化建設委員會 http://www.cca.gov.tw/	May 1998-April 2005	84	146	1.7	119	1.4
gov	行政院勞委會 http://www.cla.gov.tw/	Dec. 1998-April 2005	77	202	2.6	50	0.6
gov	行政院交通部 http://www.motc.gov.tw/	Feb. 1997-April 2005	98	170	1.7	44	0.4
gov	行政院經濟部 http://www.moea.gov.tw/	Dec. 1996-Nov. 2004	96	159	1.7	53	0.6
gov	外交部 http://www.mofa.gov.tw/	Dec. 1997-Nov. 2004	84	90	1.1	24	0.3
gov	內政部 http://www.moi.gov.tw/	Aug. 1997-Nov. 2004	88	73	0.8	27	0.3
小計			1,231	1,890	1.5	630	0.5
org	數位典藏國家型科技計畫 http://www.ndap.org.tw/	May 2002-Mar. 2005	35	53	1.5	44	1.3
org	國家衛生研究院 http://www.nhri.org.tw/	Nov. 1998-March 2005	77	87	1.1	25	0.3
org	慈林教育基金會 http://www.chilin.org.tw	March 2001-March 2005	49	42	0.9	19	0.4
org	凱達格蘭學校 http://ketagalan.org.tw/	April 2003-March 2005	24	29	1.2	22	0.9

網站類型	網站名稱	存檔起迄	月數	存檔次數 Results	平均 存檔次數 (次數/月)	更新 次數 Updates	平均 更新頻率 (次數/月)
org	公視 http://www.pts.org.tw	Dec. 1996-April 2005	101	191	1.9	54	0.5
org	中國國民黨 http://www.kmt.org.tw/	Nov. 1996-April 2005	103	155	1.5	128	1.2
org	親民黨 http://www.pfp.org.tw/	Nov. 2000-March 2005	53	105	2.0	35	0.7
org	民進黨 http://www.dpp.org.tw/	Feb. 1997-Nov. 2004	94	97	1.0	11	0.1
小計			536	759	1.4	338	0.6
com	陳水扁 http://www.presidentchen.com/	Aug. 2000-March 2005	56	84	1.5	26	0.5
小計			56	84	1.5	26	0.5
edu	國立臺灣大學 http://www.ntu.edu.tw	Jul. 1997-April 2005	94	254	2.7	83	0.9
edu	國立政治大學 http://www.nccu.edu.tw/	Nov. 1996-March 2005	101	332	3.3	135	1.3
edu	國立交通大學 http://www.nctu.edu.tw/	March 1997-March 2005	97	210	2.2	91	0.9
edu	國立清華大學 http://www.nthu.edu.tw	Dec. 1996-April 2005	101	204	2.0	24	0.2
edu	國立師範大學 http://www.ntnu.edu.tw/	Jul. 1997-April 2005	98	226	2.3	19	0.2
edu	國立成功大學 http://www.ncku.edu.tw/	Jan. 1997-March 2005	99	400	4.0	101	1.0
edu	國立中山大學 http://www.nsysu.edu.tw/	Nov. 1996-April 2005	102	303	3.0	146	1.4
edu	國立中興大學 http://www.nchu.edu.tw/	Dec. 1996-March 2005	101	239	2.4	39	0.4
edu	國立海洋大學 http://www2.ntou.edu.tw/	Jul. 1997-April 2005	94	147	1.6	28	0.3
edu	國立東華大學 http://www.ndhu.edu.tw/	Feb. 1997-March 2005	98	264	2.7	15	0.2
edu	淡江大學 http://www.tku.edu.tw/	Dec. 1997-March 2005	88	251	2.9	77	0.9
edu	文化大學 http://www.pccu.edu.tw	Feb. 1997-March 2005	98	251	2.6	148	1.5
edu	慈濟大學 http://www.tcu.edu.tw/	Dec. 1998-March 2005	76	116	1.5	19	0.3
edu	東海大學 http://www.thu.edu.tw/	Dec. 1996-Oct. 2004	95	219	2.3	87	0.9
edu	輔仁大學 http://www.fju.edu.tw/	Dec. 1996-Nov. 2004	96	131	1.4	88	0.9
edu	大同大學 http://www.ttu.edu.tw/	Jan. 2000-March 2005	63	100	1.6	41	0.7
edu	逢甲大學 http://www.fcu.edu.tw/	Dec. 1996-March 2005	101	164	1.6	33	0.3
edu	元智大學 http://www.yzu.edu.tw/	Dec. 1998-March 2005	76	139	1.8	113	1.5
edu	長庚大學 http://www.cgu.edu.tw/	Dec. 1996-March 2005	101	142	1.4	45	0.4

網站類型	網站名稱	存檔起迄	月數	存檔次數 Results	平均 存檔次數 (次數/月)	更新 次數 Updates	平均 更新頻率 (次數/月)
edu	中原大學 http://www.cycu.edu.tw/	May 1997-March 2005	95	169	1.8	14	0.1
edu	國立臺灣大學圖書館 http://www.lib.ntu.edu.tw	Jun. 1997-April 2005	95	134	1.4	81	0.9
edu	淡江大學學生紀念圖書館 http://www.lib.tku.edu.tw/	Nov. 1996-March 2005	101	100	1.0	35	0.3
edu	中華經濟研究院 http://www.cier.edu.tw/	Jan. 1998-March 2005	87	52	0.6	30	0.3
edu	行政院教育部 http://www.edu.tw/	Nov. 1996-April 2005	102	230	2.3	153	1.5
edu	中央研究院 http://www.sinica.edu.tw	Jan. 1997-March 2005	99	363	3.7	129	1.3
edu	國立自然科學博物館 http://www.nmns.edu.tw/	Aug. 1997-Mar. 2005	92	163	1.8	116	1.3
edu	國家圖書館 http://www.ncl.edu.tw	Dec. 1996-Apr. 2005	101	184	1.8	49	0.5
小計			2,551	5,487	2.2	1,939	0.8

資料來源：本文作者研究整理。

說明：本文表四係依據此查詢結果所作之統計。查詢日期：2006 年 6 月。

Web Preservation: with Introduction to Required Techniques and Applications

Shu-fen Hung Lin

Abstract

This article discusses the purposes, necessary planning and techniques for Web preservation. The author addresses the importance of preserving Internet resources both on open access and close access. By exploring some Web preservation examples of other countries, the author points out some strategies for technical practice and collaborative approach. Besides, the author works on searches of Taiwan's webpages of different kinds of institutions in the Internet Archive

Keywords (關鍵詞) : Web Preservation ; Web Archiving ; MINERVA ; PANDORA ; Internet Archive ; Kulturarw³ ; WayBack Machine ; Web Clawing ; Open Access Resources ; Close Access Resources ; Institutional Repository ; Whole Domain Model ; Comprehensive Approach ; Selective Model ; Thematic Model ; Deposit Model ; Combined Approach ; Web Information Accessibility ; Accessibility Verifying Tool ; Accessibility Software
網路資源備份保存；網路資源搜尋；網路資源可即性；網路開放性資源；網路閉鎖性資源；機構典藏；全域模式；選擇性模式；主題式模式；寄存式模式；混合式模式；網頁可即性檢核工具；網頁可即性輔助工具

Shu-fen Hung Lin : Executive Officer, Special Collections Department, National Taiwan University Library ; Doctorial Student, Graduate School of Library and Information Science, National Taiwan University, Taipei, Taiwan ; E-mail : shufen@ntu.edu.tw

WayBack Machine to find out how Taiwan's Web resources had been archived. The investigation shows that structure of a website, as well as design of webpages contained in it, is crucial for whether or not the website could be completely archived. Furthermore, the metadata information attached to a webpage is used for automated cataloging for its archival version and will indirectly influence users' search. Since a successful Web preservation task is due to good planning by both the information producer and the archival institution, the author tries to cover the issue concerning Web resources cataloging and the necessary effort by the Web information producer. It is hoped that the information and suggestions of this paper will evoke attention, solicit the enforcement of related activities so as to ensure the long-term preservation and usability of digital cultural heritages of the nation.

