大學圖書館
11 卷 2 期（民國 96 年 9 月） 頁 1-18

臺灣大學網站典藏系統之研發

The Development of National Taiwan University Web Archiving System

陳光華 **Kuang-hua Chen**
國立臺灣大學圖書資訊學系副教授兼圖書館系統資訊組組長
Associate Professor,
Department of Library and Information Science, National Taiwan University
Head, Information Technology & Services Department,
National Taiwan University Library
E-mail: khchen@ntu.edu.tw

丁培峰 **Peng-fung Ting**
國立臺灣大學圖書館計畫助理
Project Assistant, National Taiwan University Library
E-mail: jerryting@ntu.edu.tw

【摘　要】

網站典藏是一個新興的概念，主要是為了保留瞬息萬變且具文化歷史意涵的網站網頁，近幾年國外已經逐步發展且建立了一些網站典藏系統，然而對於國內的圖書館員來說，卻仍然是一個比較不熟悉的議題。臺灣大學圖書館關注網站典藏的發展已經有很長的一段時間，決定於 2006 年起著手研發網站典藏系統，期間雖然面臨不少的困難與挑戰，但是也發展了對應的策略。本文介紹國外的網站典藏計畫，再說明臺灣大學網站典藏系統的研發過程、實作細節、問題與困難，期盼分享網站典藏的技術與經驗，提供大家參考。

【Abstract】

Web archiving is an emerging concept whose main purpose is to preserve websites of cultural or historical significance. In this paper, we describe the NTU Web Archiving System, NTUWAS, that was developed by the National Taiwan University Library and launched in 2006. In order to help readers not be familiar with web archiving, we give a brief survey of current web archiving activities in the world. We then describe the implementation and functionalities of NTUWAS, as well as the difficulties and problems that we encountered when carrying out this project.

臺灣大學學術期刊資料庫

關 鍵 詞：館藏發展；數位典藏；網站典藏
Keywords：Collection Development; Digital Archive; Web Archive

一、前言

一九九三年是網際網路發展歷史的分水嶺，圖形化瀏覽器Mosaic的出現，使得網際網路的使用群從特殊領域的少數使用者轉為各種領域的各色使用者，無論是學者專家或是普羅大眾都能夠很容易地使用Mosaic，以及後來的 Netscape Navigator、Internet Explorer、Opera、Firefox、Safari 等瀏覽器。對於內容創作者（Content Creators）而言，發表成果的管道也解放了，使得網際網路上擁有數以億計的網頁，承載著各式各樣的資訊；各種類型的機構也紛紛建構網站（Websites），使之成為官方資訊的匯集中心。演變至今日，網站與其組成的網頁，已經具有多重的意義了，不僅是訊息性的，知識性的，文化性的，也是歷史性的，這些的特質造就了「網際網路」成為一種可以反應當代政治、社會、經濟、文化現象的載體。設若加入時間的訊息，前述各層面的資訊，將成為歷時性的現象描述，對於五十年、百年後欲從事這個時代研究的學者專家們，將是極具價值的歷史素材。

但是，吾人若深入思考會發現，網際網路訊息或資源的生命其實是很短暫的。十年前的舊報紙或許可以在圖書館中找到，試問十年前的電子報今日在哪裡可以發現呢？網路資訊流通地非常快速，網路資訊的保存卻成為一個燙手的問題，根據調查，平均一個網站的生命週期大概只有 44 天（Library of Congress, 2003），而重要的資訊，例如學者發佈的文獻資料、文化或科學資源這些寶貴的資訊，可能就因為時間的更迭而從此消逝掉了。再以選舉活動為例，現在與過去也有了極大的不同，紙張文宣減少了，個人競選網站與部落格（blog）反倒成了親近選民的必備工具，成為向選民宣達政績、政見、理念、行程的聯絡處；然而選戰資訊時時更新，當選戰落幕，這些網站瞬間關閉，等到研究者想鑽研相關史料時，這些第一手研究素材早已煙消雲散。

理解網站網頁內容的價值與存在的意義後，一個直接的結論應是今日若不典藏重要的網站網頁，未來將會有一大塊的時代記憶空白，學者與專家將無法回顧這個時代的發展，因為沒有原生研究材料，一般的群眾也很難去追溯這個時代的點點滴滴。

Internet Archive 的 Wayback Machine 計畫的精神標語，非常具有代表性，它是這麼說的 “Archiving the

臺灣大學學術期刊資料庫

Internet for the Future Generations”。（參見 Wayback Machine, http://www.archive.org/web/web.php）

昔日實體文獻的典藏，是圖書館的重要任務之一，並訂定具有特色的館藏發展政策。但隨著網際網路時代的崛起，保存典藏具有研究價值的網站及網路文獻，成為大學圖書館必須積極面對的責任，應是數位典藏的一環，同時也屬於館藏發展的範疇。臺灣大學圖書館對於「網站典藏」（Web Archives）這項議題，投入長期的關注與探討，於2006年開始進入系統開發的階段，並確認網站典藏的範圍與政策，致力於「選擇（select）」、「蒐集（collect）」、「保存（preserve）」臺灣大學、臺灣政府機構、臺灣政治團體、焦點人物、原住民族及重大事件之原始性網頁資料，並提供檢索服務以滿足當代與未來世代研究發展之需求，冀望這樣的開展能夠為臺灣善盡保護網路資產之責任。本文將就臺灣大學網站典藏系統（NTU Web Archive System，簡稱NTUWAS）發展之實際經驗，提供諸位先進參考，期待進一步引發彼此之交流與迴響。

本文第二節將簡要說明目前網站典藏的現況；第三節逐步說明臺灣大學圖書館的實施作為；第四節討論問題與困難；第五節則是簡短的結論。

二、網站典藏的發展現況

網站典藏的概念與發展在臺灣仍屬於剛起步的階段，反觀國外，非營利組織 Internet Archive，早在1996年於舊金山成立，致力於收藏數位格式的檔案，其目的在於提供研究者、歷史學家、學者可以永恆的存取歷史性數位資料。Internet Archive 典藏文字檔案、音訊檔案、視訊檔案、軟體系統、以及Web網站。典藏網站可說是非常具有開創性的作法，在眾人積極投入數位圖書館的研究，將重要的文件、檔案、器物、標本、畫作等等轉換為數位形式，供更多的使用者瀏覽，Internet Archive 也特別關注原生數位的 web 網站，建構 Wayback Machine 系統，進行網站典藏的工作，因為網站的版本變動是更加地頻繁。另一方面，澳洲國家圖書館 Pandora 計畫（National Library of Australia, 2007）亦從1996年開始投入網站典藏的研究領域，而後美國國會圖書館（Library of Congress, 2006）、英國、日本等國家也陸續投入類似的研究計畫。以下將簡單介紹國外推動網站典藏機制，建構網站典藏系統的的計畫。

(一)MINERVA

（http://memory.loc.gov/cocoon/minerva/）

MINERVA（Mapping the INternet Electronic Resources Virtual Archive）

臺灣大學學術期刊資料庫

是美國國會圖書館的網站保存計畫，其有主題性的深入蒐集美國網站資料，在完整蒐集主題相關的網站後，並進行細緻的分類。以選舉主題的網站為例，又細分為候選人、政府、政黨、民意等等，目前還正在蒐集 911 事件網站、2002年冬季奧林匹克運動會、伊拉克戰爭，請參考表 1。對於網站內容，MINERVA 盡可能複製一份存檔，使用的網頁蒐集工具為 Heritrix（Library of Congress, n.d.），包含網站上的 HTML 網頁、圖片、PDF、Flash、聲音、影像檔皆完整收錄，截至2007年5月，已蒐集了超過 70TB 的資料（Library of Congress, 2007），也已經收錄超過 36,000 個網站。目前並不提供檢索功能。有關MINERVA的詳細介紹可以參考 “Collecting and Preserving the Web: The Minerva Prototype” 一文。（Arms, Adkin, Ammen, & Hayes, 2001）MINERVA的首頁請參考圖1。

(二) Wayback Machine

（http://www.archive.org/web/web.php）

Wayback Machine 附屬於 Internet Archive 計畫的一部分，廣泛地蒐集網際網路上各種類型的網站，其目標是想成為獨一無二的網際網路上的網站快照儲存體。Wayback Machine 和 MINERVA 計畫相同，也特別建構主題典藏，例如南亞大地震、卡崔娜颶風和瑞塔颶風、國家檔案、911、2000年美國總統大選、以及Web Pioneer 等等主題。此外，Wayback Machine 建立使用者論壇，強化與使用者的互動，亦建構 FAQ，以回答使用者的常見問題。然而，Wayback Machine 系統功能比較陽春，目前只提

表 1　MINERVA 網站典藏主題

典藏主題	網站數	採集網站期間
第 107 屆國會	588	2002/12/12
2002 年選舉	4000	2001/7/1～2002/11/30
911 事件	30000	2001/9/11～2001/12/1
2000 年選舉	797 個	2000/8/1～2001/1/21
2002 年冬季奧運	70	2002/2/8～2002/2/24
911 回憶錄	1800	2002/9/11～2002/9/18
伊拉克戰爭	正統整中	2003/3/1 迄今
第 108 屆國會	正統整中	2004/1/1～2004/12/31
2004 年美國總統大選	正統整中	2003/10/1～2004/12/31

臺灣大學學術期刊資料庫

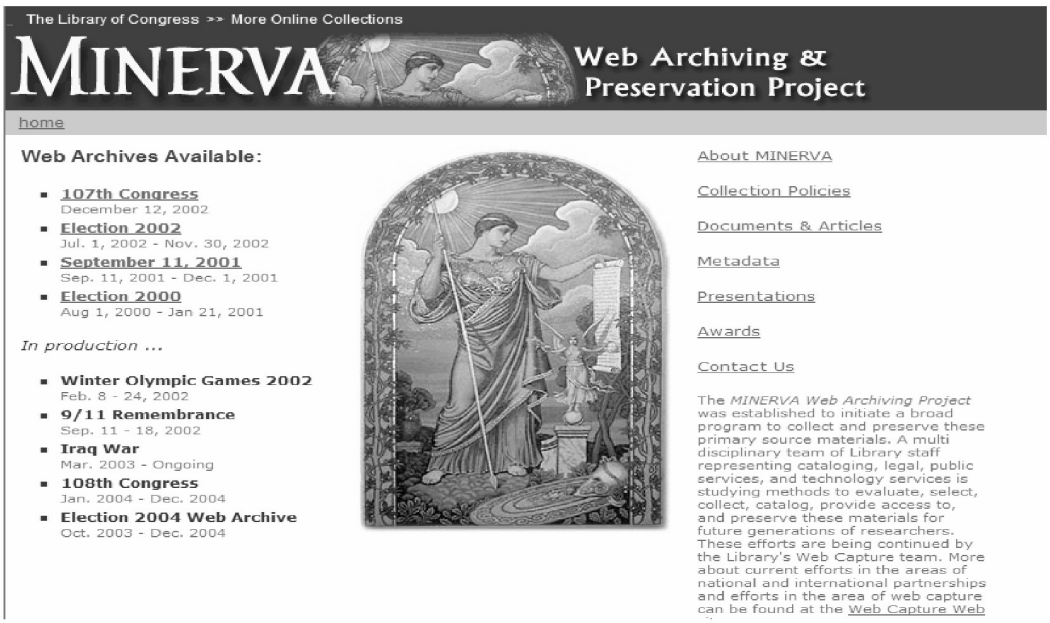


圖 1　MINERVA 網站首頁

供介於兩個日期間的網址檢索功能，不支援關鍵字檢索或是全文檢索功能。Wayback Machine從1996年開始已蒐集了將近 850 億頁，200TB 的資料，但蒐集與審核似乎較不嚴瑾，有些網站也不見得可以瀏覽。Wayback Machine的首頁請參考圖2。

(三)PANDORA

（http://pandora.nla.gov.au/index.html/）

1996年成立，最初由澳洲國家圖書館投入發展，目前則是與澳洲九所大學院校共同開發系統功能，並建立一套名為PANDAS的管理系統。該計畫主要蒐集澳洲的網站，包含社會、政治、文化、宗教、科學、經濟等議題相關的網站，共分成十五大類典藏主題，清楚的分門別類，並依照英文字母予以排序，提供使用者檢索。使用的網頁擷取工具（Internet Robots，Web Crawlers）為 HTTRACK，有關 PANDORA 計畫的技術細節，可以參考 “PANDORA: Technical Details” 一文。（McPhillips, 2004）PANDORA的首頁請參考圖3。

(四)UKWAC

（http://www.webarchive.org.uk）

UKWAC（UK Web archiving consortium）於2004年6月成立，由British Library（Lead partner）、The National

臺灣大學學術期刊資料庫

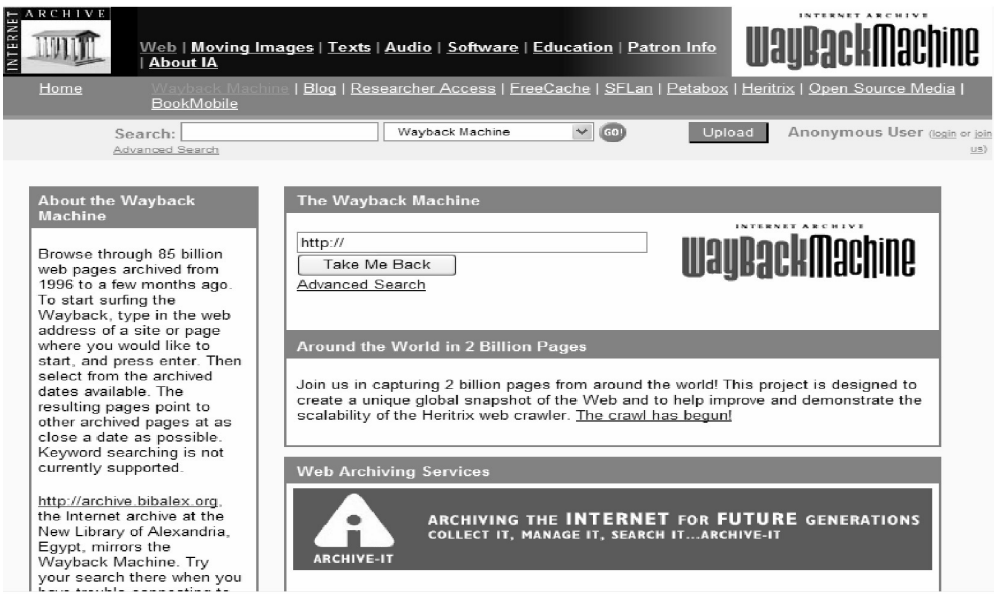


圖 2　Wayback Machine 網站首頁



圖 3　PANDORA 網站首頁

臺灣大學學術期刊資料庫

Archives、National Library of Wales、National Library of Scotland、JISC、Wellcome Trust 等六個機構，一起投入經費共同合作，研發相關的網站典藏技術。並由2005年開始有選擇性的收錄英國相關的網站，每個機構可以訂定自己的典藏政策，各有其收藏的主題。像Wellcome Library專著於醫學的網站，British Library 專注於文化歷史或重大政治議題，National Library of Wales則蒐集能夠反應現代Wales生活的相關網站。UKWAC 使用澳洲國家圖書館 PANDORA 計畫的開發工具 PANDAS，因此其系統架構與澳洲 PANDORA 相似。此外，UKWAC 在收錄網站前會先經過該網站的擁有者同意才收錄。有關 UKWAC 計畫的細節可以參閱 “UKWAC: Building the UK’s First Public Web Archive” 一文。（Bailey & Thompson, 2006）UKWAC 的首頁請參考圖4。

(五) WARP

（http://warp.ndl.go.jp/）

WARP （Web ARchiving Project），系統網站是由日本國會圖書館負責研發，2002年開始實驗性質的開辦，於2006年計畫臻於成熟，仍由日本國會圖書館負責整個計畫的執行。目前 WARP 已經收錄三千多個網站，所有上傳的網站都有詳細的審核和授權機制，可以自我推薦或經由他



圖 4　UKWAC 網站首頁

臺灣大學學術期刊資料庫

人推薦，主要收藏中央／地方政府、公營／法人／大學網站、特殊重要主題的網站，爾後會慢慢擴大到非政府機關的網站，網站目錄除依照網站特質分類，也依照圖書館分類法分類，一年一次下載完整網站，亦提供關鍵字與全文檢索功能。WARP 的首頁請參考圖5。

(六)中國 web 信息博物館（Web Infomall）

（http://www.infomall.cn/）

中國 web 信息博物館是北京大學網絡實驗室負責架構，每月平均以4,500 萬個網頁的速度擴展，截至2007年6月，已經有25億頁以中文為主的網頁。目前的系統並不會下載圖片影像，預計下一個版本才會開始蒐集圖片影像。除了蒐集網站網頁，和美國國會圖書館 MINERVA 計畫相同，也建置主題性的網站網站典藏。Web Infomall 除了自行設定典藏範圍外，也允許有意將網站網頁永久典藏的機構申請網站的上傳。目前 Web Infomall 系統僅提供網址的搜尋，必須輸入完整網址才能尋找到網站資料，對使用者較為不方便。Web Infomall的首頁請參考圖6。

前述各計畫的比較，整理如表2。整體來看，美國國會圖書館的“MINERVA” 是有主題性的典藏網頁，因此內容較為深入；“Wayback



圖 5　WARP 網站首頁

臺灣大學學術期刊資料庫



圖 6　中國 Web 信息博物館網站首頁

表 2　Web archive 網站比較表

網站名稱	國家	使用工具	特　色
MINERVA	美國	Heritrix	深入，有主題性地深入蒐集美國網站
Wayback Machine	美國	Heritrix	廣泛，蒐集全球網站
PANDORA	澳洲	HTTRACK	館際合作共同開發，分類清楚檢索容易
UKWAC	英國	HTTRACK	館際合作共同，各校蒐集各有其主題與特色，分類清楚檢索容易
WARP	日本	Wget (ver.1.8.2)	精緻，審核和授權機制詳細，採用圖書館分類法進行網站分類，網站 Help 說明清楚。
Web Infomall	中國	N/A	提供允諾網站下載的許可文件

Machine” 是廣泛的蒐集網站，因此範圍較廣泛；日本的 WARP 較為精緻，系統網站的使用說明很清楚，並有特別的分類法；澳洲的 PANDORA 與英國的 UKWAC 主要是由館際合作共同開發系統，此外參與 UKWAC 計畫的機構有各自的蒐集主題，分工較為清楚；大陸的 Web Infomall 正在持

續發展，除了廣泛蒐集網站網頁，也進行主題性的網站典藏，但是因為正在發展，因此系統功能較為陽春。

二、臺灣大學網站典藏

臺灣大學網站典藏系統（NTUWAS），主要以「前端讀者介面」，提供詳盡的原始性網站資訊，來服務讀者；輔以「後端管理系統」，由館員依照已擬定之政策來蒐集各類別網站，並進行管理；並透過「自動排程程式」來下載網站網頁；最後典藏網頁於系統儲存體（網站典藏庫），形成一完整的循環鏈結，請參考圖7。

(一)前端讀者介面

主要展示出NTUWAS典藏的網站資訊，讀者藉由搜尋與分類檢索，可以檢視典藏的網站資訊。主要功能如下所示。（參見圖8）

1. 關鍵字檢索：目前提供讀者全文檢索來搜尋網頁資料、以關鍵字來搜尋網站資料。（參見圖8的⑴）
2. 網站簡介：向讀者說明NTUWAS的緣起與目的，說明由內而外的典藏策略。（參見圖8的⑵）
3. 典藏統計：統計目前典藏的網站類別與各類別所蒐集的網站數量。讀者可藉此功能清楚地瀏覽各類別的網站資料，可深入瀏覽各網站，並且檢視同一個網站不同時期的典藏資訊。（參見圖8的⑶）
4. 最新典藏：顯示目前最新收錄的網站清單，並可檢視最新的收錄的網站資訊。（參見圖8的⑷）
5. 熱門點選：顯示目前最熱門點選的網站清單，依據點選次數排序，並可檢視該網站的相關資訊。（參見圖8的⑸）
6. 典藏縮影：顯示典藏網站的縮影圖示，提供使用者一個簡要的視覺印象。（參見圖8的⑹）

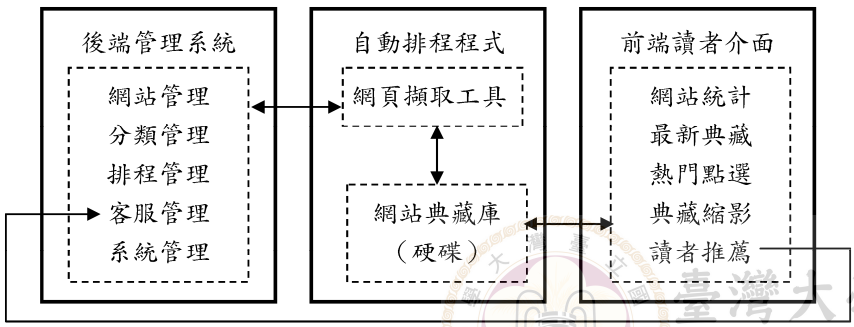


圖 7　臺灣大學網站典藏系統架構圖

臺灣大學學術期刊資料庫



圖 8　臺灣大學網站典藏系統網站首頁

7. 讀者推薦：邀請讀者一起來做網站的主人，讀者可推薦未收錄的網站，以彌補收錄的不足。（參見圖 8的(7)）

另外，首頁的「典藏統計」功能將典藏網站的不同版本做一個統整與

臺灣大學學術典藏

分類（參見圖 8 的(3)），讀者藉由點選各個類別的網站內容，可以瀏覽到各類別的典藏網站清單（請參見圖 9）。

當使用者點選「庫存網站」連結時，會顯示該網站名稱、網址，並列出過去蒐集的庫存日期清單（圖 10）。讀者可以直接點選典藏網站的庫存日期，即可調閱出當時的網站資料，如圖11即為1997年臺灣大學網站的首頁，對於多數的臺灣大學師生而言，應該都已經不記得這個網頁了，



圖 9　典藏網站清單

臺灣大學學術期刊資料庫

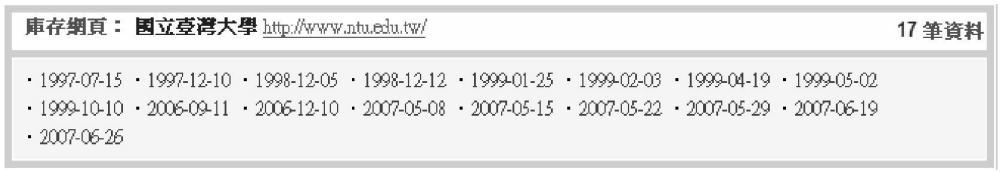


圖 10　典藏網站內容



圖 11　1997 年版之臺灣大學網站首頁

但是，這樣的系統卻可以喚起大家的記憶。

(二)後端管理系統

除了前端完善的讀者介面，尚需功能完整的後端管理系統，才能使網站資料源源不絕，建立起良好的循環。後端管理系統主要提供館員管理欲典藏的網站資料，並建立排程的相關設定，以便收錄網站與維護網站。

臺灣大學期刊資料庫

主要功能如下所示。

1. 網站管理：管理網站資料，提供館員新增、維護、刪除網站。此外，還有檢核網站資料的功能，其目的為避免在讀者面前顯示出錯誤的網頁資料，因此網站下載完成後必須先檢核網站資料，以確保資料的完整性。
2. 分類管理：管理網站之分類資料，提供管理者新增網站分類、維護網站分類、刪除網站分類。建立完畢後，後續既方便讀者區分、瀏覽網站，亦方便館員將網站分門別類。
3. 排程管理：主要提供館員設定網站的下載排程，依照網站不同的特質設定不同的下載週期，之後由自動排程程式將這些網站資料依設定的週期下載。
4. 客服管理：主要管理讀者所推薦的網站資料，並經過館員篩選以決定是收入館藏。
5. 統計分析：主要提供管理者網站相關的統計分析資料，如典藏網站數量、網站瀏覽次數、網站下載次數、網站最新瀏覽日期，以便提供管理者作為決策判斷的參考。
6. 系統管理：主要維護後端之管理功能、使用者資料、使用者權限及部門資料。

(三)自動排程程式

藉由工作同仁所擬定之收錄政策，設定網站的下載排程。定期或不定期（即時）從網際網路自動下載網站資料。

1. 定期下載：基本上，系統會依不同類型的網站，預設適當的下載週期，而工作同仁也可以依照網站不同之特性，設定不同的下載週期，例如：臺灣大學戲劇學系可能每月都會更新展演的訊息，就可以設定每月下載，而某些歷史記錄網站，可能不會再變更了，只需要收錄一次，因此可以依照不同的狀況設定網站的下載排程。「自動排程程式」就會依照這些設定，自動派遣網頁擷取程式去下載網站網頁。
2. 即時下載：有些網站可能因事件的發生而有其時效性，透過排程下載可能為時已晚。因此館員可以即時下載網站，即時典藏重要的網站，以免失去收藏的完整性。例如，近期「中」字輩的企業醞釀改名，如中華郵政、中國石油等企業，若不即時下載網站，可能原先珍貴的網站資源，從此消逝不見。

臺灣大學網站典藏系統，相較於其他類似的計畫，提供了更為詳盡的檢索功能，讓讀者可以方便的搜尋網站網頁資料，表 3 簡要整理本文提及之網站典藏計畫與NTUWAS計畫的比較，敬請參考。我們更提供「典藏縮影」可以讓使用者參考系統典藏的網站首頁。後端管理系統提供完善的管理介面，工作同仁可以謹慎的選擇網站，藉由自動排程設定，自動下載網站，亦節省了龐大的人力的成本負擔。並儘可能下載完整的網站內容，

臺灣大學圖書館術期刊資料庫

表 3　Web archive 網站功能比較表

網站名稱	全文檢索	分類檢索	網址檢索	讀者推薦	典藏縮影	主題回顧
NTUWAS	✓	✓	✓	✓	✓	*
MINERVA						✓
Wayback Machine			✓			
PANDORA	✓	✓	✓			
UKWAC	✓	✓				
WARP	✓	✓		✓		
Web Infomall						✓

*：NTUWAS 亦有許多主題型的網站典藏，如九二一大地震、二二八事件等。

完整地保存，完整地呈現在讀者面前。

四、研發重點與問題討論

NTUWAS發展的過程中，遭遇到了不少的困難，也發現了發展網站典藏的不少重點，在此將技術細節與所遇到的困難分享出來，提供大家參考。

(一)研發重點

1. 展現不同版本的網站網頁：由於網站下載完成後，存在於一個個的資料夾內，如何讓網站資料下載到 Server 上，並使讀者透過網路能夠查詢，檢視到同一個網站不同時間的典藏歷程，然後將兩者聯繫起來，讓讀者能瀏覽到典藏的網站，實是一個難題。為解決此一問題，我們有系統的為每一個下載排程命名，讓每一個網站的每一個下載排程都有一個資料夾可以互相對應。爾後，當讀者查詢網站時，就可以跟典藏的網站連結上，而管理者也能清楚的分辨尋找出典藏的網站資料，增進管理的效能。
2. 提供全文檢索功能：很多國外網站典藏系統網站不見得提供全文檢索的功能，當初考量為提供讀者更好的服務，因此結合 Java Lucene，提供全文檢索功能。
3. 使用功能完善之下載工具：NTUWAS 使用 HTTRACK 作為擷取網頁的工具，HTTRACK 提供完整的下載功能，能將網站完整的下載下來保存，讓管理者無後顧之憂。
4. 擬定準則用心典藏：網際網路涵蓋的範圍極為龐大，必須訂定網站典藏準則，否則將會無窮無盡而漫無邊際。因此，臺灣大學網站典藏系統擬定典藏準則，針對符合準則的網站予以典藏，以建立具有主題特色的網站典藏。

臺灣大學圖書館 期刊資料庫

5. 建立審核機制去蕪存菁：臺灣大學網站典藏系統的網站通常藉由館員蒐羅與讀者推薦而來，透過審核機制，篩選符合收藏準則網站予以收錄，收錄下載後並經過檢測，確定網站內容無誤才提供給讀者檢索，以確保典藏的品質。
6. 建置獨立運轉的服務系統：建議分別建置擷取網站資料的伺服器與提供服務的伺服器，以免彼此干擾影響系統效能。一般而言，網頁下載工具程式（如 HTTRACK）耗費較多的電腦系統資源，因此必須建置獨立運轉的使用者服務系統，才能發揮最大化的效益。
7. 指定專人管理：網站典藏是一個需要長久發展的數位典藏新領域，而且屬於館藏發展的範疇，但是因為網站網頁極為龐雜，需要有專人用心來管理維護，才能確保網站典藏系統的永續發展。

(二)問題討論

1. FLASH 網頁難以解析：由於 Flash 的網頁連結，鑲嵌在 Flash 程式之內，因此無從知道網頁的連結，因此就算藉由HTTRACK下載Flash內容，可能其所連結的網頁也無法下載下來，造成下載的不完成，這個可說是目前世界各國網站典藏計畫亟欲探討的問題。
2. 動態連結難以臆測：由於某些網頁的連結是由Javascript程式碼動態產生，因此當下載工具去抓取網頁時，可能無法抓取到連結的資訊，因此造成網頁資料遺漏，由於Javascript 程式的寫法不一，並無一定的準則可以作為標準來判斷。
3. 典藏空間需求龐大：網站典藏系統需要旁大的儲存空間，會耗費很大的成本與資源，如 Internet Archive 計畫已經蒐集了超過200T的網站網頁，不可能完整蒐集所有的網站。因此，網站典藏系統需要有完整的目標規劃，並考量該如何節省空間資源，以謀求永續的發展，否則漫無目的的收錄網站，將會佔用過多的資源與成本。
4. 網站編目尚無定論：目前圖書館的圖書編目，使用MARC格式且有一定的編目規則。然而，網站網頁應該使用何種 Metadata Format，應該如何編目，國際間尚無一定之準則，應該要發展網站網頁之編目方法，這也是極具研究性的議題。
5. 圖書與資訊的結合：網站典藏屬於數位典藏的一環，對於圖書館來說是一個新的領域，也是不能不重視的一個新領域（項潔、洪筱盈，2006），更需要圖書資訊與資訊科技的專業結合才能激盪出燦爛的火花。習慣於圖書領域的圖書館館員，面對於此新領域的典藏工作，難免會有認知的差距，而資訊科技的工作者又不熟悉圖書領域之專業知識，這兩者間該如何合作，才能架構出完美的網站典藏系統，也是有待努力的課題。

臺灣大學圖書館期刊資料庫

五、結論

網際網路時代，資訊習以數位的型態呈現，未來將會更日益頻繁，日常的書報、雜誌、與新聞亦有可能同時發表紙本與數位版本，因此，圖書館為善盡保存文化資產之責任，將網站典藏視為數位典藏的一環，同時也屬於館藏發展的範疇，是有其必要性。然而，網路範疇無限廣大，不可能無限量的典藏各類型的網站，所以，圖書館必須制定出符合自我需求的典藏準則，由內而外地，先保存校內單位、系所本身的網站、學術研究或研討會議網站，其次是典藏與校園社區相關的網站，再擴及到其他範疇。另外還需依據網站不同之特質，訂定不同的更新週期，以每天、每週、每月或每年為週期來典藏網站。圖書館亦可依其校園特色發展獨立典藏之主題，例如臺灣大學圖書館另設有「臺灣原住民族圖書資訊中心」，因此，NTUWAS特將原住民議題，納入典藏之範疇，藉以強化館藏特色。

網站典藏的建置與發展包含有歷史性、文化性、政策性、技術性的課題，需要有不同專業的人士共同合作，才能畢其功。臺灣在網際網路的發展並不遜於其他國家，臺灣在20世紀末期以迄於今的發展又極具時空的特殊性。若能由國內各個圖書館的團結合作與分工，建立具有各自特色的網站典藏系統，再建置聯合目錄之機制，互通有無、由點到面，完整呈現臺灣在各種領域的網站網頁，這樣的館際合作模式將會有助於臺灣文化資產的典藏與維護。

致謝

感謝臺灣大學圖書館NTUWAS計畫團隊成員項潔館長、張素娟主任、郭美玲編審、阮紹薇編審、何庭輝先生、黃于鳴小姐、廖偉辰先生的共同合作。

參考文獻

Arms, W. Y., Adkins, R., Ammen, C. & Hayes, A. (2001). Collecting and Preserving the Web: The Minerva Prototype. RLG DigiNews, 5(2). Retrieved June 20, 2007, from http://www.rlg.org/preserv/diginews/diginews5-2.html.

Bailey, S. & Thompson, D. (2006). UKWAC: Building the UK's First Public Web Archive. D-Lib Magazine, 12(1). Retrieved June 26, 2007, from http://www.dlib.org/dlib/january06/thompson/01thompson.html.

Internet Archive (n.d.). About the Internet Archive. Retrieved June 26, 2007, from http://www.archive.org/about/about.php.

臺灣大學學術期刊資料庫

Library of Congress (2003). Web Capture & Archiving. Retrieved June 20, 2007, from http://www.loc.gov/acq/devpol/webarchive.html.

Library of Congress (2006). About MINERVA. Retrieved June 26, 2007, from http://memory.loc.gov/cocoon/minerva/html/minerva-home.html.

Library of Congress (2007). Web Capture: FAQs. Retrieved June 26, 2007, from http://www.loc.gov/webcapture/faq.html.

Library of Congress (n.d.). Technical Background. Retrieved June 26, 2007, from http://www.loc.gov/webcapture/technical.html.

McPhillips, S. (2004). PANDORA: Technical Details. Retrieved June 26, 2007, from http://pandora.nla.gov.au/pandoratech.html

National Library of Australia (2007). About Pandora. Retrieved June 26, 2007, from http://www.nla.gov.au/.

項潔、洪筱盈（2006）。大學圖書館數位保存與館藏發展策略：以臺大圖書館為例。大學圖書館，10(2)，頁 2-18。

臺灣大學學術期刊資料庫