網路資源長期保存：以多元層次描述模式建構之探討

王麗蕉

編審
中央研究院台灣史研究所
博士生
台灣大學圖書資訊學研究所
E-mail: lcwang@gate.sinica.edu.tw

摘要

本文主要目的在於探討網站資源多元層次組織架構，以及Web資源長期保存的多元層次描述之可行模式。首先介紹網際網路、全球資訊網的發展與Web資源組織標準；其次，列舉國際間相關Web資源保存計畫，並分析各保存計畫所採用的資訊組織規範；闡述檔案來源原則之理論基礎與控制層次應用，並分析Web多元層次組織架構；據以探討Web資源多元層次描述（multilevel description）之可行模式；最後，提出關於Web資源長期保存發展之建議。

關鍵詞： 多元層次描述模式，Web資源組織規範，檔案編排，來源原則，Web資源長期保存

前　言

「保存人類知識、傳承人類文明」是圖書館自古以來一直堅守的偉大使命。在資訊網路環境快速發展，全球資訊網興起與普及化後，諸如紙張等傳統載體經過數位化而產生之再生與利用（再生數位），不僅各式各樣原生數位之電子資源蓬勃發展，特別是全球資訊網（World Wide Web，簡稱Web）資源的成長更形巨大。圖書館界與資訊界不再爭論資訊的擁有與使用孰輕孰重，轉向共同研究數位資源長久保存與傳承人類知識文明而努力，各國國家級知識機構開始進行Web 保存計畫。鑑於數位資源之長久保存，不僅要提供現時使用，更需要考量人類文明的延續，以及提供未來世代知識的使用。資訊編排與描述等組織行

2007/04/26投稿; 2007/05/28修訂; 2007/06/13接受

為，是資訊利用的基礎，如何對各式數位資源作有效與一致性的描述，在Web環境下更顯其重要性。以主題內容為選擇性之Web保存，是所謂主題導向的蒐集，相關典藏資源涉及蒐集者主觀判斷的可能偏差，將資源抽離原有情境的其他連結資源，可能導致遺漏更重要的資源與研究價值等缺失。

檔案編排之來源原則，尊重資源產生者及其原有資源結構，具體實施的控制層次，有別於主題內容分類的主觀性，具有組織工具的客觀性與目的合理性，又能呈現資源產生的背景情境，提供資源歷史價值的運用。Web資源在資訊網路環境，由網站、首頁與網頁組合而成，網頁資訊內容可包含文字、圖形、聲音、影像、視覺資訊或其組合；為方便網頁資訊使用與管理，將網頁資訊分門別類，再以超連結方法連接相關內容的網頁，使個別網站所有資訊成為一個整體。網站資源則是具有結構化的架構，以來源原則與控制層次為基礎編排與描述，是基於領域的基礎，具客觀與合理性。本文主要目的在於探討網站資源多元層次組織架構，以及Web資源長期保存的多元層次描述之可行模式。

本文首先介紹網際網路、全球資訊網的發展與Web資源組織標準；其次，列舉國際間相關Web資源保存計畫，並分析各保存計畫所採用的資訊組織規範；接著，闡述檔案來源原則之理論基礎與控制層次應用，以及分析Web多元層次組織架構；並據以探討Web資源多元層次描述（multilevel description）之可行模式；最後，提出關於Web資源長期保存發展之建議。

二、資訊網路環境發展與Web資源組織規範

網際網路（Internet）與Web可謂為20世紀資訊界最偉大的發明，使人類知識文明邁入資訊網路時代。Internet的起源，肇始於1950年代美蘇冷戰下的軍事競爭，自1969年採用Paul Barran的「封包交換」理論，發展出「網路控制協定」（Network Control Protocol, NCP），將UCLA等四所大學的電腦主機連結成單一網路—ARPANET，即為Internet的前身；至1980年改採「TCP/IP」通訊協定，奠定現在Internet的基礎（註1）。並於1984年制定「網域名稱轉換機制」與「網域名稱伺服器」（Domain Name Server, DNS），提供易記、便利的連接機制，進而成就日後Internet蓬勃發展（註2）。Internet發展歷程及其相關資訊科技的發明與運用，參見圖1。

1993年Web正式進入Internet之服務，結合Internet原有之電子郵件（E-mail）、遠端連線（Remote login），與檔案傳輸（File transfer）三大主要功能，成為人類生活不可或缺的資訊傳播管道。Web是以圖形界面方式，結合Internet上多種既有資訊傳輸協定的新興網路技術，支援文字、聲音、圖像與視覺等多樣性資源，以超連結方式，組合成全球化資訊網路服務。

Web資訊空間，主要是由網站（Web site）、首頁（Homepage）與網頁（Web page）所組成。所謂網站係指「Internet上特定空間，以一全球資源定址器（Uni-

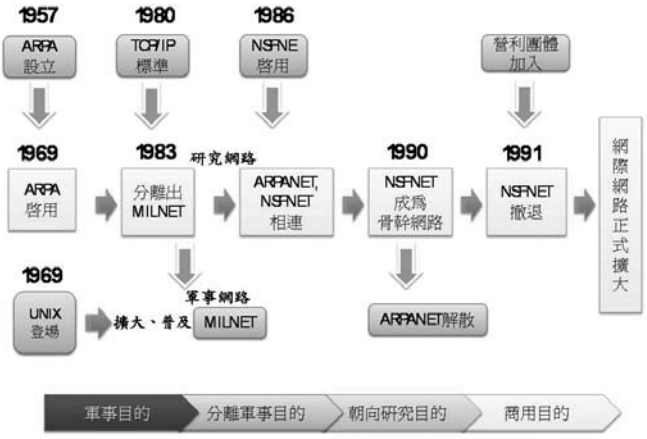

圖1　網際網路（Internet）發展歷程

資料來源：依據台灣網路資訊中心、PC Office，*My Name! My Site! My Style*（台北市：台灣網路資訊中心，2004），6-7整理。

form Resource Locator，簡稱URL）連結使用，包含文字、影像與其他資料形式，任何人都可透過Internet取用。特定網站是由URL中主機名稱（hostname）加以辨識」。一個網站通常有一首頁，通常就是其主機名稱，例如，http://www.loc.gov/（註3）。因此，網站是指網際網路上有特定網域，提供超媒體、超文本及超連結的網路服務。每一網站伺服器都有一個獨一無二的IP位址與相對應的網域名稱（Domain name），網站中提供多媒體形式的網頁資源，網頁資訊內容可包含文字、圖形、聲音、影像、視覺資訊或其組合；為方便網頁資訊使用與管理，將網頁資訊分門別類，再以超連結方法連接相關內容的網頁，使個別網站所有資訊成為一個整體。

Web資源具有超文本、超媒體與超連結等特質，對於Web資源之組織，為因應Web資訊網路環境，圖書館界在評估MARC標準的適用性與擴展性後，開始著手研究更合適的資訊組織方式與標準，國際間相關知識機構亦發展出各式各樣不同用途的詮釋資料（metadata）標準，目前適用於文字性電子資源之描述性詮釋資料有DC（Dublin Core）與EAD（Encoded Archival Description）。此外，英國相關數位圖書館計畫所發展的CLD（Collection Level Description），亦受到英國重要數位計畫所採用，如博物館暨圖書館暨檔案館委員會（Museums, Libraries, and Archives Council, MLA）的Cornucopia，以及英國國家檔案館的A2A計畫等。前述三種詮釋資料的發展及其規範的元素，茲簡述如下。

㈠Dublin Core(DC)

成立於1995年的Dublin Core Metadata Initiative組織，目的在於：1.發展metadata標準，以利各領域資源的探索；2.制定各式metadata互通架構；3.促進

各社群或特定學科metadata資源發掘、交換與共享。在2003年DC已成為ISO標準，其強調Item層次的使用，主要以item為描述單元。目前最新版本共有主題與關鍵字、題名、著者、相關敘述、出版者、其他參與者、出版日期、資源類型、資料格式、資源識別代號、關連、來源、語文、內容範圍、版權規範等15項元素組成（註4）。

㈡Encoded Archival Description(EAD)

起源於1993年美國柏克萊大學的查檢工具計畫，至1998年制定第1版，並由美國國會圖書館負責維護，又於2002年公布第2版。EAD重視資源多元層次結構關係，以全宗（record group或collection）為最高層次，在全宗下列出資源結構及其描述性資訊，例如Franklin Roosevelt廣播演講，所有紀錄可依日期條例相關書目性資訊列在其collection level之下（註5）。其主要描述資訊為辨識性描述、行政資訊、編排、傳記與歷史、檢索控制、附註、附屬資料描述、其他描述資料、組織、範圍與內容，以及附屬組成描述等11大群組。

㈢Collection Level Description(CLD)

英國相關數位典藏發展組織在開發數位圖書館計畫時提出有關Collection Level Description之需求，1999年由英國圖書館網路辦公室（UK Office for Library Networking, UKOLN）在JISC的研究支援圖書館計畫（Research Support Libraries Programme, RSLP）贊助下，在2000年完成「RSLP Collection Description Schema」文件，規範館藏層次描述（Collection Level Description, CLD），描述項目大致可分一般性描述、主題、日期、相關機構、外部關係等項目（註6）。

相較於圖書館界通用的機讀編目格式（MAchine-Readable Cataloging, MARC），DC是規範較精簡的描述項目，EAD是以可群集的描述項目，對於網路上數位資源的描述皆比MARC好用。但要找到一個放諸四海皆準的標準則不可能，因此，資源的整合與互通，不同詮釋資料的對映研究亦陸續發表，例如，DC對映到MARC，EAD與MARC的對映等。

三、網路資源長期保存計畫及其資源組織現況

自20世紀中葉以來，網際網路的發展與Web興起等資訊科技的重大變革，人類知識傳播主要媒介已經由傳統紙張轉移到數位形式，而Web正是數位資源傳播主要的網路環境，目前，Web 可謂是世界上最龐大的數位資源集中地。有鑑於Web資源已成為人類知識主要形式，而Web資源快速成長及迅速消失的特性使國際上開始注意到Web site resources保存對人類知識傳承的重要性，遂開始進行Web資源保存的相關計畫。包括1996年起，澳洲國家圖書館（National

Library of Australia, NLA）所進行的Pandora計畫，美國舊金山非營利組織推動的Internet Archives，以及美國國會圖書館（Library of Congress, LC）於2000年開始著手的Minerva計畫。

網站資源保存的蒐集方式主要是以「網站快照」（Snapshots of Web Sites）下載方式一蒐集與保存的單元為一個網站，以電腦結構的觀點是檔案的組成。網站蒐集技術則是利用程式從Web伺服器複製檔案到一台電腦之上，定期執行下載快照，因此檔案館能擁有每一網站循序的快照（註7）。然而下載快照的方式，仍有些實務上的問題，包括：

㈠格式：網站中的檔案格式具多樣性，諸如：文字、影像、視聽等，包括以JavaScript或Java語言設計的程式和相關表單或metadata檔案。有些網頁不斷產生或依據資料庫中的資訊，並每年更新不同格式，或新版格式不斷出現。

㈡個別網站的界限較難定義：通常以網站中具有某一相同網域名稱或相同URL的所有資料以作為蒐集範圍。

㈢網站的檔案有很多錯誤或不一致現象：通常不去確認格式或不存在的超連結，鏡錄程式會加以辨識或紀錄在異動檔中。

㈣時間的安排：下載整個網站通常需要耗費數個小時，若其間出現資源更新的狀況，導致快照容易出現不完整或發生不一致的現象。

㈤資料庫：下載網站資源主要是以儲存在Web伺服器的資源；在網站中提供使用介面的資料，通常很難為保存目的而進行下載，除非得到網站出版者的同意。

㈥取用性：網站下載失敗，可能導因於電腦故障或網路斷線。因此，區分檔案的暫失性或無法使用，是否因檔案已被移除或不曾存在是無法達成的任務。

㈦系統效能：網站資源的鏡錄程式，在相關電腦與網路設備的配合方面，至少要四小時，因此任何抓取程式都要考量效能與價值的平衡（註8）。

蒐集策略主要有二種：㈠Bulk蒐集：所有符合標準的網站都加以蒐集，例如Internet Archive的政策是蒐集所有HTML網頁。㈡Selective蒐集：由圖書館員或其他專家針對個別網站評選，澳洲的Pandora計畫的蒐集政策，由圖書館員依澳洲特定興趣主題加以選擇，並決定每一網站蒐集的頻率（註9）。

目前國際間重要的網站資源保存計畫有非營利組織的Internet Archive、澳洲國家圖書館的Pandora與美國國會圖書館的Minerva等。茲就其發展背景、蒐集與典藏方法，以及資源組織方式，分述如下。

㈠Internet Archive – Universal access to human knowledge

圖書館的存在意義是保存社會文化的產物並提供使用，在數位科技時代，

圖書館若要繼續扮演教育與學術角色，必須發展圖書館在數位環境中的功能。若沒有文化產物，人類將沒有學習成功記取失敗的記憶與機制，同樣的，若隨Internet的爆炸，將會生活在如同Danny Hillis所謂的「數位黑暗時代」。因此，Internet Archive企圖為研究人員、歷史學家與學者提供數位形式之歷史館藏的永久使用。成立於1996年美國舊金山，Internet Archive是一非營利組織，自1996年起開始定期蒐集開放取用（Open access）的HTML網頁，約每個月一次，至2001年3月，五年間已擁有60兆位元資料，每月成長10兆位元。由Alexa Internet之商業公司負責蒐集資料，六個月後再捐贈給Internet Archive（註10）。

目前與美國國會圖書館等機構共同合作，蒐集與典藏的內容包含Web資源、動態影像、文字、視聽資源與軟體等五大類。最終目的在於將人類知識提供全球化之取用（Universal access to human knowledge）。在Web資源的蒐集與保存方式，是以整個網站內容的快照，定期蒐集。以網站主機的URL為蒐集與保存單元，依定期抓取時間條列在其URL之下。提供Wayback Machine（時光機器）的使用介面，自1996年迄今（2007年）已保存550億網頁，可輸入網站位址瀏覽不定期抓取之快照資源內容。但目前尚未支援關鍵字的查詢（註11）。目前，Wayback Machine提供重要的Web Collection有：Katrina和Rita颶風、國家檔案館（英國中央政府網頁檔案館，是一選擇性典藏的英國政府網站）、美國2002選舉、September 11th（2001年的911事件）、美國2000總統大選，以及Web Pioneers（Web先鋒）等專輯館藏。

㈡Pandora – Preserving and Accessing Networked Documentary Resources of Australia

澳洲國家圖書館進行的Pandora計畫，可稱為Web資源保存計畫的先鋒之一。起源於1996年，主要目的在於建置一個蒐集網路資源之澳洲線上出版品的檔案館，並發展長久保存的國家性策略（註12）。目前是由NLA與其他九個澳洲圖書館與文化典藏組織共同合作建置。計畫名稱PANDORA 之由來，依據其計畫目的「保存與使用澳洲網路文獻資源」（Preserving and Accessing Networked Documentary Resources of Australia）。期望在資訊網路時代，繼續發揮典藏國家文獻傳統的偉大使命。至2005年，NLA認為Pandora檔案館，是典藏評選性澳洲線上出版品的世界級檔案館，保存具研究或文化意義的電子期刊、政府出版品與網站資源（註13）。

澳洲國家圖書館對於「線上出版品」（Online publication）的類別主要區分為政府出版品、澳洲網站快照、商業出版品、地圖、音樂、資料庫、電子期刊、會議論文、線上遊戲、部落格、入口網站、新聞網站等。經過長保存價值、永久使用性、資訊技術、人力資源與經費成本等考量，Pandora收錄的線上出版品包括：1.公開的政府出版品，2.教育機構出版品，3.會議論文，4.電

子期刊，5.索摘代理商提供的item（通常為前述四種類別，但具有紙本版本），與6.指定主題領域營運三年以上和記載現時社會、政治事件的主要議題的網站，如選舉網站、雪梨奧運、巴里島爆炸等。不收錄的類別有：1.資料庫，2.線上報紙，3.新聞網站，4.討論群、通報和新聞群組，5.部落格（支援學術出版品除外），6.入口網站，與7.線上遊戲（註14）。其中關於網站資源的蒐集政策，Pandora計畫是採選擇性策略，NLA持續評估整個澳洲網域定期快照的可行性，亦曾考量與Internet Archive合作，但由於IA只有蒐集到部分澳洲網域資源，對個別出版品仍未達到完整的蒐集；以及考量整個澳洲網域典藏的技術複雜度，目前NLA仍以選擇性方式企圖達成：

1. 每一蒐集的項目（item）都品質鑑定及現有技術支援最大使用功能保證；
2. 檔案館的item是完整編目，並納入國家書目；
3. 檔案館中的item都經由與出版者協商取得公開使用的保證；
4. 資源的重要優先性是為增進保存需求的知識而分析與決定（註15）。

網站選擇的標準在於考量其主題內容。為支援大量資料的徵集與管理，以及協助合作伙伴以遠端連線工作站進行更有效率保存典藏之建置，NLA於2001年正式推出Pandora Digital Arching System（PANDAS），並持續發展軟體功能。

相關網站資源的編排與描述，如同其他線上出版品或電子資源的處理標準，參考*AARC2*, 2nd ed., 2002 revision, Ch.9「電子資源」之內容標準，與MARC結構標準，以及NLA的「電子資源編目手冊」（*Electronic Resources Cataloguing Manual*）。網站編排與描述單元，或以單一網站之整合性資源，或以單一Collection為主體。NLA對「Collection」的定義是一電子資源的群組，分享共同主題或基於一個事件，在檔案館中以collection呈現，以增進使用者資源查詢與檔案館的管理（註16）。相關編目基本原則：

1. 為協助Kinetica（澳洲國家書目網）與OPAC查詢有用資訊，包括所有Australian Internet sites的資源，所有蒐集的條目題名皆以方括表示，如[Sports - Australian Internet sites]，便於使用者查詢。
2. 編製多個足以包含collection廣泛特性的主題標目，網站中所有個人與機構則以附加款目表現。
3. 必要時，提供有關新網站增加的蒐集時間起迄的註記，和個人或機構的檢索款目。
4. 當新的網站加入一個Collection時，更新每一相關個人或組織之附加款目（註17）。

Pandora網站首頁，提供藝術與人文、商業與經濟、電腦與網際網路、教育、環境、健康、歷史與地理、原住民、青少年、法律與犯罪學、新聞與媒體、政治與政府、科學與技術、社會與文化、運動與休閒等15項主題瀏覽，以及資源查詢服務。

㈢Minerva – Mapping the Internet Electronic Resources Virtual Archive

國會圖書館是美國最古老的聯邦文化機構，肩負保存知識與提供資源使用的歷史使命。進入21世紀資訊網路時代，LC開始思考「原生數位」資源的取得與永久使用等相關議題。於2000年起開始進行Minerva Prototype計畫，目的在於增進有關蒐集與組織選擇性網站的實務議題的討論，以及探究LC應如何進行全面性保存計畫。LC曾請求Internet Archive提供有關於2000年秋季總統大選的網站，直到2001年就職典禮，提供約150到200個網站每天的抓取（註18）。計畫名稱「MINERVA」是根據其對映網際網路電子資源虛擬檔案館（Mapping the Internet Electronic Resources Virtual Archive）之目標而來的。Minerva蒐集政策是以選擇性方式蒐集，是針對主題內容而不論其格式或類型。選擇標準有：1.符合現在或未來國會與研究人員之資訊需求；2.唯一、獨特資訊；3.學術內容；4.遺失的風險（主要在於網站資源短暫性的特質）；5.資訊的時效性（註19）。

Minerva保存資源描述方式：以主題（如，2000大選）蒐集網站資源，並以*AACR2*與MARC建置Collection level的目錄紀錄，以便於在圖書館整合系統（Integrated Library System）提供查詢。由於每一主題包含數千個網站，因此研擬（metadata Object Description Schema, MODS）標準，使能在Collection下建立title-level的描述（註20）。截至2003年4月，蒐集的網站已超過3.6萬個。目前Minerva網站上，提供2000大選、911事件、2002冬季奧運、2002選舉與第107屆國會等主題典藏的查詢服務。

總括而言，非營利組織Internet Archive、澳洲國家圖書館的Pandora以及美國國會圖書館的Minerva等Web資源保存計畫，就其Web資源組織而言，Internet Archive以大量、整個網域快照蒐集方式，所典藏的資源並未進行整理與描述，因此僅能提供URL查詢、瀏覽網頁，而無法提供關鍵字或更深入的內容查詢服務；Pandora主要考量網頁主題內容加以選擇蒐集，以整個網站，或collection等二種描述單元，參考*AARC2*與MARC等有關電子資源編目標準，除了建置Pandora查詢系統外，並使所蒐集的網頁資源之編目紀錄能納入其國家書目網，以提供一致性的書目查詢服務；Minerva是以主題選擇方式蒐集與保存網頁資源，亦以*AACR2*與MARC作為collection層級的編目標準，但在collection下採用MODS標準作為title層級的描述標準，提供二元層次的整理與描述。

四、檔案來源理論及其應用於Web資源保存模式

Web資源之長久保存與開放使用，其核心在於網頁資源的編排與描述方式。Web資源如同檔案資料，是一有機成長的彙集性資源，資源之間具有相

互連結等結構性組織。檔案界通常依賴的來源原則編排理論與控制層次描述模式，其主要特徵為：㈠檔案的描述通常著重於全宗，全宗是由一組文獻組合而成，並包含不同形式媒體；㈡尊重檔案來源（Provenance）；並包含複雜的階層結構，以漸進式分析，從全宗到個別文件；㈢強調檔案資料的智能結構與內容等（註21）。來源原則之編排依據，有別於主題內容選擇的主觀性，具有客觀、理性基礎，符合Web資源有機成長的特質；檔案控制層次漸進式描述模式，對於具彙集性的Web資源，是一具有效率與效益的可行模式。

有關檔案來源原則的理論基礎，以及應用多元控制層次編排Web資源的可行模式，茲分述如下。

㈠檔案編排來源原則的理論基礎

檔案編排來源原則之現代化檔案科學發展，最初起源1841年法國檔案學者Natalis de Wailly所提出關於「尊重全宗」（respect des fonds）之概念。不久，普魯士（現德國）汲取法國檔案館之尊重全宗原則，於1881年頒佈普魯士國家機密檔案館檔案整理條例提出了登記室原則與尊重原始順序原則（respect for original order）。至1898年由S. Muller、J. A. Feth和R. Fruin三位荷蘭檔案學者在其*Manual for the Arrangement and Description of Archives*（簡稱為「荷蘭手冊」）一書，正式定義為「來源原則」，並加以完整闡述。來源原則之概念，具體實施方式為控制層次（Level of control），其最佳闡釋是由Oliver Wendell Holmes於1964年所提「現代檔案工作重點是由廣泛與一般性到最細微與特定性，以漸進方式，彙集與描述檔案單元」（註22）。

檔案相關來源原則、尊重全宗、尊重原始順序，以及控制層次等原則與方法，作為檔案編排與描述之基礎架構與內涵說明如下：

1. 概念思想：來源原則

檔案編排之來源原則，在於檔案是隨著機構或個人之業務與活動所產生之文書，是呈有機成長，並經過有條件的價值鑑定，才得以成為長久保存的檔案資料，而經由檔案可反映出某一機構或個人的特質，因此檔案編排必須依據其來源，以了解檔案的出處（註23）。來源原則提供檔案人員與研究者能正確了解檔案產生的來源與目的。T. R. Schellenberg歸納來源原則之檔案編排，其優點有：⑴可維護檔案的證據價值；⑵符合檔案的特質；⑶可協助檔案人員處理檔案；⑷便於檔案的編排；⑸易於檔案的描述（註24）。

2. 具體表現：尊重全宗

檔案來源原則的核心在於檔案的產生單位，強調檔案與機構的關係，即以檔案產生之有機體—行政機構、家庭或個人，作為檔案編排整理的單位。Terry Cook提出五點判定全宗的標準：法定的地位或法律名稱（a legal identity）、官方命令、行政階層中的位置、一定程度的自主權，與一個組織圖（註25）。

3. 內化發展：尊重原始順序

尊重原始順序原則是用以維護政府機構歸檔系統，包括特殊的文書系列與其相互之間的關係。「荷蘭手冊」亦提到「基於檔案館藏的原始組織，主要可符合其產生機構的組織架構」（註26）。運用尊重原始順序處理檔案在於原始順序具有下列益處：⑴反映當時業務的確實情況；⑵保存文件原有的關係；⑶提供有關記錄產生、利用或活動的文書證明；⑷增加檔案價值，例如提供其文書活動的公正證據（註27）。

4. 實施方式：檔案控制層次

來源原則實際應用的控制層次主要區分為全宗、系統、案卷與件等四大基本層次，從整體性到特定性，以漸進方式，彙集與描述檔案單元。

現代檔案科學發展，檔案編排以來源原則之理性基礎的概念思想，具體表現於外是尊重全宗，內部延伸結構為尊重原始順序，實際應用方法為檔案控制層次，構成了理論與實際兼具的專業科學，其理論性架構，參見圖2。

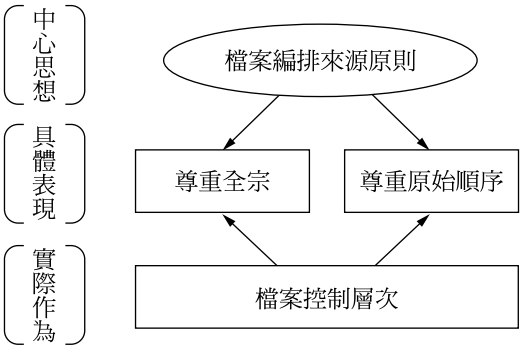

圖2 檔案編排來源原則的理論架構關係

來源原則之具體實施是以檔案控制層次方式進行，在全宗、系列、案卷與件等四大主要層次，依尊重全宗與尊重原始順序，由一般性到特定性，從最廣泛到細微處，以層次漸進方式，符合目的理性的實踐。四大基本層次中，每一控制層次都有其行政、處理需求和檢索等內涵與資訊。

1. 全宗（Record group）：「全宗」一詞源於法文fonds，英美國家則是以archive group或record group代表檔案全宗，手稿資料則以collection表示手稿全宗。通常由一機構文書或一個人文件組成，檔案人員應蒐集該全宗檔案的一般性內容與其整體的歷史或傳記資訊（註28）。

2. 系列（Series）：系列是在全宗之下，依機構的下屬單位、業務，或功能的記錄組成，包含有產生文書之特定業務與歸檔架構等資訊。由於檔案有彙集性與相互關連性等特質，故以系列作為檔案描述最基本的單元。其描述項目應包含：題名、日期、檔案數量、實體編排、內容摘要等（註29）。

3. 案卷（File folders）：系列下包括許多案卷，主要是檔案產生時，為便於管理與保存，透過立卷依一定順序或標準，將性質相同的文件集合，稱為案卷。案卷編排方式有：⑴字母順序，⑵年代順序，⑶地理區域，⑷主題，⑸數字等五種標準（註30）。

4. 件（Item）：是指個別文件，是案卷下的組成單元；案卷下的件，可能是一件公文、一封書信、一張照片等，通常依日期或字母等一定順序排列。

檔案編排來源原則、尊重全宗、尊重原始順序等三原則，具體實施之控制層次，提供全宗、系列、案卷與個別文件等四項基本控制層次，可形成多元層次編排（Multilevel arrangement）之基礎，以建置多元層次描述（Multilevel description，MLD），達成多元層級描述資訊，提供深入的檔案內容查詢與使用之服務。

㈡Web資源之多元層次組織架構

來源原則之編排依據，有別於主題內容選擇的主觀性，具有客觀、理性基礎，適合Web資源有機成長與彙集性結構之特質。在全球資訊網空間，Web資源是以「一致性資源定位器」（URL），即網址，以定義好的格式，用於呼叫網路上各式各樣資源，例如http、ftp、gopher、news與mailto等通訊協定的資源。URL是用來界定資源物件的位置與該物件的存取方式，由通訊協定（http、ftp、gopher），與hostname、path、filename組合而成，例如http://www.ntu.edu.tw/info/。其中hostname，即網域名稱是共享一個共同通訊位址的網路電腦群組。網域名稱是組織、企業或個人在網路上身份的代表。網域名稱的命名具有結構性，例如www.ntu.edu.tw，即為台灣大學全球資訊網。網域名稱為Internet服務中最基礎的一環，提供機器名稱與IP位址雙向對應的機制，且網域名稱比IP容易記，並具代表意義；使用網域名稱讓系統更具移植性，當IP變動時，只須更改網域名稱伺服器（Domain Name Server，DNS）設定即可（註31）。

DNS於1984年制定了第一個規範（RFC1034、RFC1035），介紹網域名稱概念及其階層次屬性的概念。DNS整體架構為一樹狀結構，經由全球唯一的Root Server達到正確搜尋的目的。DNS之樹狀結構，每一分支以「.」分隔，限制：最多127層、每個分支最多63字元（A-Z，0-9，-），總長為256字元。網域名稱採階層式管理：網域名稱的管理者可建立不同的子網域給不同的部門單位使用，亦可將此子網域授權他部門自行管理，在上一層的網域須指出這種授權的關係（註32）。

因此，Web資源編排可依據網域名稱之階層式結構，進行組織網站資源的控制層次，Web資源多元層次的界定與組成，簡述如下。

1. 最高層：機構或個人網站

將Web資源的產生機構視為一個整體，以機構的網域名稱作為第一層，網

域名稱是組織、企業或個人在網路環境中身份的代表，可作為定義一個全宗之明確的界定。例如：www.ntu.edu.tw 代表台灣大學的網域名稱。

2. 第二層：網站下子網域或主要內容與服務項目

在機構網域下，可根據網站服務或功能，區分成不同Web系列。例如：Info.ntu.edu.tw台灣大學學生資訊網。

3. 第三層：在子網域或服務項目下相同性質的網頁

在Web系列下，由於Web資源更新便利的有機成長，為便於保存與管理，可經由彙集相關性質的網頁，依日期或字母順序等次序加以編排。

4. 第四層：個別網頁

網頁是Web資源的最基本組成元素，有文字檔、影像、聲音、動畫等資源形式，在網路空間是以URL界定與存取。

依網域名稱與URL結構所界定的Web資源控制層次，是尊重網站產生者—機構或個人之原始結構加以定義，如同檔案來源原則與控制層次的具體實現，具備了客觀性且符合目的理性的理論基礎。二者的對映與Web資源多元層次應用，參見表1。

表1 Web資源多元組織架構

	檔案控制層次	Web資源多元層次組織
最高層	檔案全宗（或手稿合集）	機構或個人網站網域名稱
第二層	系列	網站下的附屬單位或服務項目
第三層	案卷	內容或服務下，相同性質網頁
最底層	個別文件	個別網頁

以來源原則為理論基礎，依據檔案控制層次應用方式，所建構之Web資源多元階層結構，除以網域名稱為最高層級的來源，以了解網站產生者之機構歷史與成立宗旨外，運用個別網站之網域結構與尊重網站內容歸類順序，更能符合Web資源產生背景與內容結構。

前述Web資源長期保存計畫中，澳洲國家圖書館的Pandora計畫與美國國會書館的Minerva計畫，都以Web collection作為主要描述層級，主要以特定主題或重大事件為主的「主題導向」（Subject-oriented）所蒐集與典藏之collection，著重其內容的描述性metadata。而依據Web資源產生機構的來源原則，則是基於領域（Domain-based）的理性基礎，以機構網站之網域名稱及其DNS之階層結構，進行多元層次編排，除了可提供內容（content）描述外，更可針對Web資源的關連（relation）與背景脈絡（context）提供完整、詳細的描述資訊。

由尊重來源原則所建構之多元層次描述模式，由整體性到特定性，從最廣泛到細微處，以循序漸進方式，符合Web資源有機成長的彙集性與結構性，達

到Web資源組織之工具客觀與目的合理。多元層次描述應用於Web資源組織之架構如圖3所示，每一Web資源之控層次都有其管理、保存與檢索等需求之不同資訊內涵。

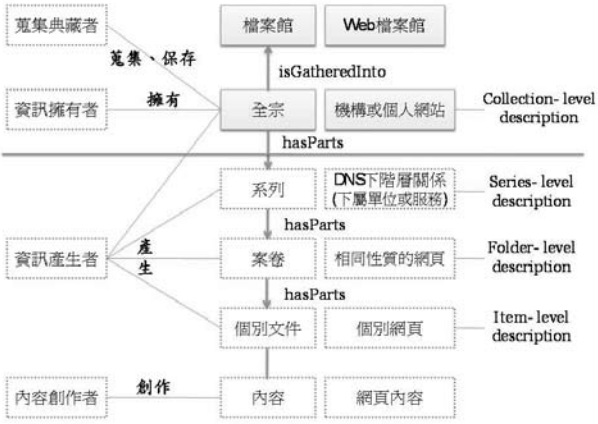

圖3　Web資源多元層次描述模式

各個層次描述的資訊內容深度和其層級是成反比的，即最高層級的描述內容是簡要概述，最底層的描述是深入至個別網頁內容。茲就各層次描述內容，分述如下：

1. **最高層：機構或個人網站**

此是依據檔案來源原則的理性基礎，尊重Web資源的產生單位，即網站產生者—機構或個人。有別於Pandora和Minerva所定義collection之主題內容導向，而以機構或個人網站為一個整體來源，是以網域為基礎的客觀性架構。因此，描述項目包括網站的產生者、擁有者與蒐集保存者等三方面的資訊需求，主要描述資訊應有網站產生者背景資訊—機構歷史或個人傳記、網站內容概述、網站取用規範、網站保存需求、網站組織管理等，對Web保存管理者而言，最重要是蒐集網站產生機構或個人的歷史描述和整體網站的內容概述。

2. **第二層：網站內子網域或服務項目**

機構或個人網站之下屬子網域或服務項目，通常是網站產生者就內容主題性與服務功能性加以分類的項目，第二層描述項目應包含單元名稱、內容摘要、主題、資源涵蓋日期、資源組織等。

3. **第三層：相同性質的網頁**

在第二層子網域或服務項目下，由網站產生者將具有相同性質的網頁，加以彙集呈現，以方便資源的管理與使用。第三層描述項目主要為單元名稱、資源涵蓋日期、資源格式等。

4. **最底層：個別網頁**

是Web資源最基本的元素，是指單一個別網頁，在網路空間中以URL定義，提供連結與取用，資源的格式可以是文字、影像、聲音、動畫等。對個別網頁的描述，可深入至網頁內容，包括全文等。

依據網站產生者尊重來源的多元層次組織架構，由最高層的全宗描述到最底層之個別網頁內容描述，循序漸進，以多元層次描述（MLD）模式所建置的Web資源多元層次目錄，除了提供整體性到深入性的資源描述內容資訊外，最高層次整體網站網域之描述，對於網站來源機構的歷史性描述與整體網站內容結構關係的概述，更可作為Web資源長期保存的管理基礎。

五、結　論

在Internet與Web等資訊科技日新月異，Web環境的原生數位資源已成為人類知識傳播的主流。Web資源的超文本、超媒體，與超連結的功能，伴隨的易變性與短暫生命的本質，使圖書館面臨Web資源長期保存的困難，以整個網站快照方式抓取與典藏，最能完整保存網站資源，但對資訊技術與資源成本是一項複雜且沈重的工作。因此，澳洲國家圖書館的Pandora與美國國會圖書館的Minerva等Web資源保存計畫，皆以考量內容主題或特殊事件之主題選擇性蒐集方式，但其彙集典藏的Web Collection是主題內容導向，有人為主觀的價值判斷。且抽離網頁資源原有的來源與結構，亦可能遺漏關連性資源與背景脈絡資訊等重要研究價值。

檔案來源原則的理性基礎與控制層次的具體方式，符合Web資源有機成長的彙集性與結構性，以整個網站為蒐集對象，是基於領域，尊重網站產生者的理性基礎。藉由網域名稱與URL的系統化結構，分層組織網站中所有網頁資源，並以多元層次描述模式，由上而下（top down）、循序漸進，首要工作是先完成最高層整體網站的描述，包括網站產生機構或個人之歷史性紀錄、整體網站內容概述、網站取用規範、網站內部組織與資源格式等，除呈現整體網站的產生背景與內容概述外，更可作為評量進一步分析與描述各層次的必要性與優先順序，提供資源保存者對資訊組織工作提供合理與客觀的評估標準。

面對Web資源的數量快速成長、種類日新月異、依賴軟硬體，與短暫生命週期，國家圖書館或國家級文獻典藏機構，應與各級文化典藏共同組織關於台灣網域之網站資源長期保存的合作聯盟，以「分建、共享」機制，共同建置台灣網站資源檔案館，以典藏台灣文化之原生數位資源，並提供長久保存與取用服務。

註　釋

註 1　台灣網路資訊中心、PC OFFICE，*My Name! My Site! My Style*（台北市：台灣網路資訊中心，2004），7-8。

註 2　同上註，8。

註 3　Library of Congress, "Collections Policy Statements," *Web Site Capture and Archiving* (April 2003), http://www.loc.gov/acq/devpol/webarchiv.html (accessed June 28, 2006).

註 4　International Organization for Standardization, "*Information and Documentation*: *The Dublin Core Metadata Element Set*," ISO 15836: 2003(E), (February 2003), http://www.niso.org/international/SC4/n515.pdf (accessed May 21, 2006).

註 5　Michael Seadle, "Sound Practice: A Report of the Best Practices for Digital Sound Meeting," *RLG DigiNews* 5, no.2 (April 2001), http://www.rlg.org/preserv/diginews/diginews5-2.html#feature3 (accessed June 28, 2006).

註 6　UKOLN, "CLD Online Tutorial," http://www.ukoln.ac.uk/cd-focus/cdfocus-tutorial/intro.html (accessed May 12, 2006).

註 7　William Y. Arms, Roger Adkins, and Cassy Ammen, "Collecting and Preserving the Web: The Minerva Prototype," *RLG DigiNews* 5, no.2 (April 2001), http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1 (accessed June 28, 2006).

註 8　Ibid.

註 9　Ibid.

註 10　Ibid.

註 11　Internet Archive, "Wayback Machine," http://www.archive.org/web/web.php (accessed April 17, 2007).

註 12　William Y. Arms, Roger Adkins, and Cassy Ammen, "Collecting and Preserving the Web: The Minerva Prototype."

註 13　National Library of Australia and Partners, "History and Achievements," http://pandora.nla.gov.au/historyachievements.html (accessed April 17, 2007).

註 14　Margaret Phillips, "Collecting Australian Online Publications," *Balanced Scorecard Initiative* 49 (May 2003): 3.

註 15　Ibid.

註 16　National Library of Australia and Partners, "Cataloguing Manual: Table of Contents," http://pandora.nla.gov.au/manual/kincat.html (accessed April 18, 2007).

註 17　National Library of Australia and Partners, "Preserving and Accessing Networked Documentary Resources of Australia: General Procedures," http://pandora.nla.gov.au/manual/collections.html (accessed June 10, 2006).

註 18　William Y. Arms, Roger Adkins, and Cassy Ammen, "Collecting and Preserving the Web: Minerva Prototype."

註 19　Library of Congress, "Collections Policy Statements."

註 20　Ibid.

註 21　D. V. Pitti, "Encoded Archival Description: An Introduction and Overview," *D-Lib Magazine* 5, no.11 (November 1999), http://www.dlib.org/dlib/november99/11pitti.html (accessed April 18, 2007).

註22 Fredric M. Miller, *Arranging and Describing Archives and Manuscripts* (Chicago: SAA, 1990), 28.

註23 Ibid., 19-20.

註24 T. R. Schellenberg, *The Management of Archives* (Washington, DC: NARA, 1998), 90-95.

註25 薛理桂，檔案學理論（台北市：文華，2002)，123-124。

註26 Fredric M. Miller, *Arranging and Describing Archives and Manuscripts*, 20.

註27 Ibid., 26-27.

註28 Ibid., 28.

註29 G. S. Hunter, *Developing and Maintaining Practical Archives* (New York: Neal-Schuman, 1997), 102.

註30 Fredric M. Miller, *Arranging and Describing Archives and Manuscripts*, 83-85.

註31 邵喻美，「網域名稱教育訓練：基礎篇」，台灣網路資訊中心(TWNIC)，http://www.twnic.net.tw/newdn/help/images/twnic01.pdf (檢索於2006年6月16日)。

註32 同上註。

Web Archives: The Concept and Application of Multi-level Description Model

Li-Chiao Wang

Executive Officer
Institute of Taiwan History, Academia Sinica

PhD Student
Department of Library & Information Science, National Taiwan University
Taipei, R.O.C.
E-mail: lcwang@gate.sinica.edu.tw

Abstract

With the development of Internet, Web resources have grown up rapidly and many developed countries are doing extensive researches on Web resources for a long-term preservation. The purpose of this paper is to discuss the way builds and constructs for a long-term preservation of Web resources with the multi-level description. This paper reviews the theoretical foundation of the Principle of Provenance and practical application of control level. Besides, Web multilevel framework is analyzed in order to discuss the possibility of the application of Web resource with Web multilevel description. Finally, suggestions for future research are made, analyzing in more detail about keeping development of Web resources for a long-term preservation.

Keywords: *Multilevel description; Web resources; Resources organization; Principle of provenance; Long-term preservation*