

## 第五章 美國國會圖書館數位保存計畫

本章探討美國國家圖書館—國會圖書館(Library of Congress, 簡稱LC)的數位保存的理念與計畫成果，對於美國國會圖書館的數位保存進行瞭解，不僅宏觀地檢視其國家數位保存計畫的意涵、目標與發展、法定寄存制度的發展，並微觀地瞭解美國網路資源保存計畫的發展與實施，探討數位典藏與永久典藏的功能，最後從數位保存計畫發展與策略歸納促成數位保存的因素。

本章共分五節，第一節國會圖書館與法定寄存；第二節數位保存意涵與計畫與背景；第三節數位典藏；第四節永久取用；第五節從歷史與成功策略歸納促成因素。

### 第一節 國會圖書館與法定寄存

美國國會圖書館現位於美國首府華盛頓哥倫比亞特區(Washington, D.C.)，如圖 5-1，成立於 1800 年 4 月 24 日。(註1)



圖 5-1 美國 50 州行政區域圖

其建立可追溯自美國政府成立之初，1774 年起，政府開始利用「費城圖書館公司」的資料作為釐訂政令與推展政務的參考。1789 年國會成立後，議員利用與國會同一建築物內的「紐約學會圖書館」(New York Society Library)的資料作為立法的參考，但不久就發生許多爭端，於是乃有成立國會圖書館之提議。1800 年，美國總統約翰·亞當斯(John Adams)簽署〈遷徙

法案〉，將聯邦政府遷徙至華盛頓，並根據該法案的第五條成立國會圖書館。1802年，美國總統湯瑪斯·傑佛遜(Thomas Jefferson)簽署了一條與國會圖書館有關的法案，其重要內容包括：國會圖書館的館長由總統任命；成立國會圖書館聯合委員會，負責監督國會圖書館的作業與建立適當的規則；國會圖書館的服務限於總統，副總統及國會議員。1812年起，最高法院的法官開始享有使用國會圖書館的權利，並開始借書給經館長特許的一般民眾。(註2)

由上述可知，最初是美國國會圖書館的成立主要服務的對象為美國國會，而其館藏書目都是由美國總統傑佛遜所提供。然而，後來國會圖書館被賦予了法定書庫的角色，用以保護著作權，因此所有希望得到美國著作權法保護的作者，都必須將自己作品複製兩個複本交與國會圖書館典藏。雖然現已不強制執行，但在美國出版的書籍的複本還是依例被送達國會圖書館典藏。迄今，國會圖書館收藏了許多重要圖書及收藏品，以1億3451萬冊的館藏量成為世界最大的圖書館，並奠定其成為國家圖書館之地位。

美國國會圖書館原只是單純提供立法參考服務的議會圖書館，後來逐漸發展成為全國學術界、圖書館界與一般民眾的國家圖書館，兼具議會圖書館與國家圖書館雙重角色。美國國會圖書館的使命在於使其典藏的資源能夠被國會以及美國公民使用，並擔負起支援與保存全球文化與知識之責，提供下一代創新思考力與學習能力。在行政體系與圖書館界的地位隸屬於國會，故較為能夠受惠於國會通過各種相關立法案，進行擴展規模計劃，爭取到相關經費；此外，其館長職缺係由美國總統根據〈公開法案92-310〉(Public Law 92-310)任命之，再由參議院行使同意權，其乃美國三所國家圖書館，即美國國會圖書館、美國國家農業圖書館(National Agricultural Library，簡稱NAL)與美國國家醫學圖書館(National Library of Medicine，簡稱NLM)中，唯一直接由美國總統任命者，規模也較大於其他兩所國家圖書館。

美國國會圖書館成立時僅有圖書740冊，現已成為全世界館藏資料最多的圖書館，根據2006年美國國會圖書館館藏統計，全部館藏共超過134,517,714件資料，其中包括20,532,692件於分類系統中已編目的書籍，圖書、手冊、連續性出版品、技術報告與其他印刷資料佔11,591,309冊；未分類(特殊)館藏有102,393,713件，包括：聲音資料(如磁碟、磁帶、有聲書與其他記錄性格式的資料)2,834,692件，手稿59,469,053件，地圖

4,807,827 張，微縮資料 14,047,798 件，音樂資料 5,275,222 件，視聽資料 14,299,103 件(包括動態圖片 1,023,909 張、照片 12,476,824 張、海報 92,960 張、以及圖畫 542,315 件)等。(註3) 參見表 5-1。

表 5-1 美國國會圖書館 2006 年之館藏量統計

	資料類型	數量 (冊/件/張)
已分類編目 20,532,692 件	圖書、手冊、連續性出版品、技術報告與其他印刷資料	11,591,309
	其他	8,941,383
未分類(特殊)館藏 102,393,713 件	聲音資料(如磁碟、磁帶、有聲書與其他記錄性格式的資料)	2,834,692
	手稿	59,469,053
	地圖	4,807,827
	微縮資料	14,047,798
	音樂資料	5,275,222
	視聽資料	14,299,103
總計：134,517,714 件		
視聽資料		
	動態圖片	1,023,909
	照片	12,476,824
	海報	92,960
	圖畫	542,315
	其他	163,095
總計：14,299,103 件		

國會圖書館組織龐大、體系複雜，人員眾多，多至 4,000 位館員。主要分為：(1)國會研究服務部(Congressional Research Services，簡稱 CRS)、(2)著作權局(Copyright Office)、(3)館員辦公室(Office of the Librarian)、(4)法律圖書館(Law Library)、(5)圖書館服務組(Library Service)與(6)策略計畫辦公室(Office of Strategic Initiatives，簡稱 OSI) 六大部門。

**國會研究服務部**負責提供國會立法所需的各種資料及參考諮詢服務透過即時、客觀、無黨派與機密的立法程序，提供國會立法所需的各種資料、研究、分析與資訊服務；**著作權局**為美國著作權法的執行機構，掌管各種圖書資料之著作權登記，主要任務在於藉由行政支援國家著作權系統

來提倡創新力，管理著作權法以及寄存活動，也提供國會發展國家與國際政策的建議，包括與數位保存相關的著作權問題。因此，其在 NDIIPP 計畫中扮演著重要的角色，該辦公室所提出的〈2004-2008 年策略計畫〉中，曾經提到要透過立法與修正寄存案，以增加各種不同格式數位物件徵集為目標。(註4) **國會法律圖書館**蒐藏美國以外的其他各國的法律資料，並為國會提供 32 種不同語言的法律諮詢服務，以及有關研究和法律上的各種資訊，如同聯邦法院與行政機構提供大眾法律相關的參考服務。**圖書館服務組**則是負責國會圖書館館藏的徵集、組織、取用、維護、安全性與保存工作。**館員辦公室**則是國會圖書館任務的主要責任單位。**策略計畫辦公室**負責所有與圖書館及國家數位文化資產長期保存計畫相關的策略計畫，其下設有**資訊科技服務組 (Information Technology Services, 簡稱 ITS)**掌管電腦儲存、電信通訊、安全、備份以及任何軟硬體運作相關之事宜與各項相關的基礎建設，支援服務單位及其館員的科技需求。參見圖 5-2 美國國會圖書館六大部門組織圖。美國國會圖書館組織與各部門名稱圖則詳見附錄四、美國國會圖書館組織圖。

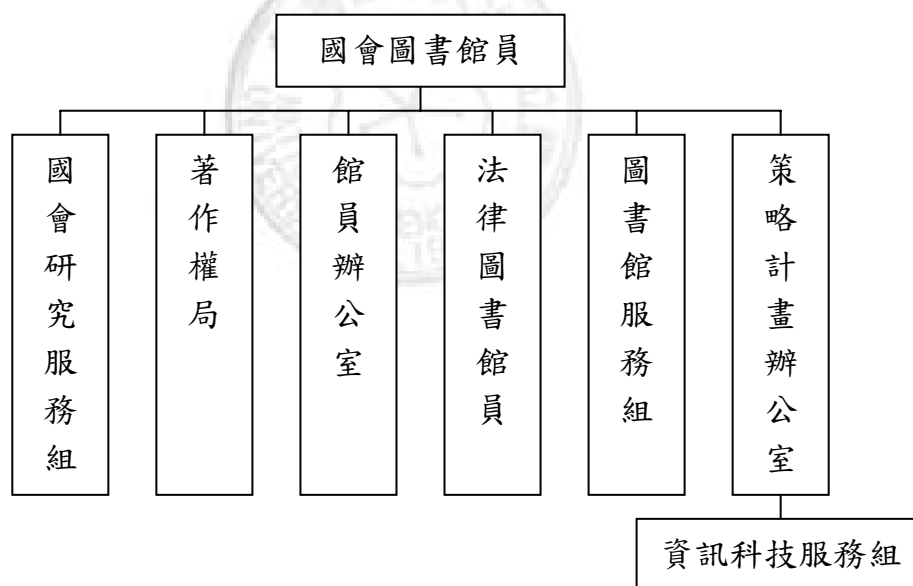


圖 5-2 美國國會圖書館六大部門組織圖

※資料來源：Ingeborg Verheul. Networking for Digital Preservation: Current Practice in 15 National Libraries. (Munich: K.G. Saur, 2006). p.244.

美國國會圖書館的性質甚為特別，除了是國會圖書館，亦是國家圖書館。因為具有國會圖書館之身分，故設立國會研究服務部，維持一個法律圖書館；又因為具有國家圖書館之身分，所以也負責全國出版物之著作權

註冊管理（註5）。

美國的法定寄存制度源自其著作權(Copyright)，其定義「著作權」是一種法律的表現形式，用於保護作者原創作品的著作權。受到法律保護的作品包括：文學作品、戲劇作品、音樂劇、藝術作品和某些其他智慧財產。此外，著作權法亦同時保障已出版與未出版的作品。在美國，規範著作權的法律主要有：1790年著作權法、1909年著作權法、1976年著作權法、1998年著作權期間延長法案(Sonny Bono Copyright Term Extension Act)、1998年千禧年數位著作權法(Digital Millennium Copyright Act，簡稱DMCA)、2005年家庭娛樂和著作權法。(註6)

因此，第一部美國聯邦著作權法最早可追溯於西元1790年，其受到英國1709年的《安娜女王法令》影響極大，現行美國著作權法則是對1909年著作權法作全面修訂，於1976年頒佈，1978年1月1日生效的著作權法，記載於《美國法典》(United States Code)第17編中，總共有8章73條。此形成美國圖書館的法定寄存制度，藉此而完整典藏美國出版品。

美國1790年著作權法僅保障作者出版「地圖、圖表和書籍」的專權。1870年，美國國會通過一項法案，使著作權系統集中由國會圖書館執行；1897年，國會圖書館開始負責著作權登記與註冊。爾後，1909年著作權法規定凡受到著作權法保護的作者，在作品出版以後3個月內，必須義務寄存兩本作品之複本，呈繳國會圖書館予以永久典藏。(註7)

1976年著作權法則允許圖書館在符合某些條件的情況下，為了保存或替換的目的可以重製著作，第106節〈Exclusive rights in copyrighted works〉規定著作權擁有者或授權他人執行下列六項義務：(1)複製作品複本或唱盤(phonorecords)；(2)根據該作品籌備衍生作品(To prepare derivative works based upon the work)；(3)貢獻作品的複本或唱盤予大眾，透過販賣、所有權的轉換、或出租等方式；(4)履行作品公開的原則，作品包括：文學作品、音樂劇、戲劇作品、芭蕾舞編舞(Choreographic)作品、手勢語言、動作片與其他音訊作品；(5)公開展示作品；(6)聲音紀錄則透過數位聲音傳送的方式履行作品公開的原則。(註8)

美國國會圖書館透過著作權法，有責任保存作品最好的版本至少一種以上的形式，擔負法定寄存之責。但是面臨數位時代來臨，數位保存日益

重要，美國對於線上出版品或是資料庫並未訂定保存版本的管理規則而無法依法使線上出版品的作者自動呈繳其作品予以圖書館典藏，是以國會圖書館在徵集數位資源以為世代永久典藏與取用時，面臨到一個典藏機關與創作者間寄存共識的無形屏障。因此，國會圖書館著作權局乃著手進行有關於電子著作權註冊和寄存的先導計畫。隨後於 1988 年，國會圖書館開始要求寄存磁帶、光碟和微縮資料。1993 年，建立了 CD-ROMs 的自願寄存協議。2001 年，第一本電子書透過電子著作權法而開始寄存，但至今仍無有關於強制或義務寄存線上電子出版品的寄存法與規定。(註9) 近來受到網際網路影響，美國國會圖書館開始探討數位保存。

## 第二節 數位保存意涵與計畫背景

### 一、數位保存意涵

科技的便捷改變了人類紀錄與展現知識的方式，人類文化智識遺產大量地以數位形式產生與儲存，有些資料甚至沒有實體形式或紙本的儲存格式，因此很有可能隨時會面臨數位內容遺失的風險，卻來不及加以典藏保存。如此一來，使用者便無法檢索與取用以數位形式出版的重要資訊資源，對於國家與社會大眾而言無疑是一種損失，乃因為此，數位保存工作更加重要與急切，美國國會圖書館認為其數位保存主要目的即在保存人類的文化知識遺產，使其能夠被國會、全國人民、甚至全世界所取用，並支援研究發展、教學，為世代子孫保留人類豐富的知識資源寶藏。

美國國會圖書館主張網路資源與數位資源保存十分重要，數位內容的保存與檢索取用是重要探討的議題，人們將生活的紀錄與出版形式慢慢地轉而由個人電腦、網路、電子郵件、數位照片等方式紀錄與呈現，使得數位內容的保存與長期檢索十分複雜。因此，數位保存必須考量到軟體與硬體的一致性、長期儲存、檔案的組織（用以提供方便搜尋與檢索）、媒體品質、災害預防、以及原生數據資料的完整性等。

國會圖書館發現過去探討數位保存，像是獨立宣言、月球登錄的電視轉播等主題，都還是根據實體媒體進行實體的保存與維護，現在開始關心的是沒有實體形式的數位資源應該如何保存？例如：網站、電子郵件、地理資訊系統產生的資料（非實體地圖）、數位化方式紀錄的音樂或視訊、代表個人資訊的網站（如：錄影資料或照片）等。(註10) 這些網路與數位

資源的典藏與保存也應該要重視。因此，國會圖書館自 2000 年乃開始探討國家數位保存計畫，特別關注於原生數位資料(Born-Digital Materials)的長期保存與永久取用。發展國家數位保存計畫與網站典藏計畫，最主要的仍是確使傳遞歷史與人類文化遺產的數位資訊能夠提供使用者永久取用之服務。

## 二、數位保存計畫與背景

數位內容的保存逐漸成為社會上重要的挑戰，因此 1998 年國會圖書館將資深管理人組成監督小組，由「聯合圖書館員策略計畫小組」(Associate Librarian for Strategic Initiatives)主持會議，發展國家數位策略。監督小組舉行多次規劃型會議，評估數位典藏與保存的現況，國會圖書館並指配「國家數位策略諮詢委員會」(National Digital Strategy Advisory Board)指引圖書館與其伙伴發展策略與規劃。2000 年，國會圖書館 James H. Billington 提出《LC 21：國會圖書館數位策略》(LC 21: A Digital Strategy for the Library of Congress)報告，建議圖書館與其他聯邦或非聯邦機構合作，並領導發展數位資訊之典藏與保存。2001 年策略計畫辦公室 (Office of Strategic Initiatives，簡稱 OSI) 成立，負責管理國家數位策略計畫各項相關事宜。

2000 年 12 月，國會贊助約 1 億元 (9980 萬) 委託國會圖書館規劃國家數位策略，發展「國家數位資訊基礎建設與保存計畫」(National Digital Information Infrastructure and Preservation Program，簡稱 NDIIPP)。由國會圖書館擔任數位策略計畫的角色與責任之因，不只因為其使命在於支援與保存全球知識館藏提供未來創新能力之外，更是因為國會圖書館是網路資源高品質內容主要的提供者之一。因此，國會圖書館便成為數位資訊領域之先導。事實上，在網際網路資源發達之前，1990-1994 年美國將其館藏選擇與數位化成為電子形式的資源，該計畫稱為「美國記憶」(American Memory)，是一個受到各界好評的 CD-ROM 計畫，後來又發展國家數位圖書館計畫，奠定日後國家數位策略計畫之基礎。(註11)

除了發展國家數位策略計畫之外，美國國會圖書館瞭解網路資料 (Web-based materials) 的文化遺產具有價值，因此積極探索各種蒐集、保存與提供網路資料的方法。於是發展「網際網路電子資源虛擬典藏計畫」(Mapping the INternet Electronic Resources Virtual Archive，簡稱 MINERVA)，建置以美國重要的主題式事件為主之網站資源蒐集與典藏計

畫，例如：2000、2002 與 2004 年的國家選舉、伊拉克戰爭以及 2001 年 9 月 11 日發生的恐怖份子炸彈攻擊雙子星大廈事件等。此計畫後來由國會圖書館 2004 年組成的網站擷取小組(Web Capture Team)負責管理。(註12)

國會圖書館在網站典藏的活動上，不僅在圖書館內部組成網站擷取小組，積極進行網站典藏的計畫與工作，更為了延續網站典藏的活動，投入於國際網際網路資源的保存活動，並支援兩所大學的網站資源典藏計畫，其一為加州大學數位圖書館的 Web at Risk 計畫，著重於保存國家政治文化遺產；其二為伊利諾大學香檳校區的 ECHO REpository 計畫，發展網路資源評鑑工具（稱為 Web Archives Workbench），測試所蒐集的網站內容之儲存、維護與取用等問題。此外，國會圖書館積極參與國際網際網路保存聯盟（International Internet Preservation Consortium，簡稱 IIPC），從全球網際網路上蒐集與保存更多、更廣泛的網路資源；以及與網際網路典藏（Internet Archive，簡稱 IA）合作軟體工具的發展。

2004 年「網站擷取小組」成立，更進一步地促進圖書館內部的網站典藏活動與技術基礎建設，協助建立圖書館界瞭解網站典藏議題。2005 年，「網站內容擷取選擇與管理」（Selecting and Managing Content Captured from the Web，簡稱 SMCCW）先導計畫成立，參考來自於不同圖書館部門的館員、管理者(Curator)、專家之建議，擴大蒐集網站的範圍，增加更多主題事件以蒐集更多相關的網路資源。SMCCW 共有 25 位館員為會員，分成四組負責不同的網站典藏計畫，包括：(1)達夫(Darfu)地區危機網站資源擷取計畫；(2)從紙本與照片延伸及加強視覺資料；(3)從現有圖書館贊助者處（包括：公民權利與政治擁護小組(Civil Rights and Political Advocacy Groups)、專業與名譽組織、研究與教育組織等）蒐集網站以延伸手稿館藏；以及(4)發展網站類型與現有的主題館藏不相關，但卻值得典藏的處理程序。(註13)

由上述可知，國會圖書館除了在國家數位資訊基礎建設與保存計畫中不遺餘力之外，亦關注網站擷取與典藏保存的計畫。以下將以圖 5-3 美國國會圖書館與數位保存計畫關連圖，呈現美國國會圖書館與國家數位保存之 NDIIPP 計畫和網站資源典藏之 MINERVA 計畫間關係，並說明兩計畫之目的與發展歷史。

在美國，數位保存計畫是以國會圖書館為主導的 NDIIPP 計畫為主，



2000 年發展的 NDIIPP 計畫，有：(1)辨別與選擇具風險性的原生數位內容；(2)建立數位保存伙伴網；(3)發展數位保存工具與服務；(4)支援公開數位保存政策；(5)宣傳數位保存的重要觀念等五項目標。在辨別與選擇具有風險性而亟需要保存的原生數位內容方面，選擇六種重要的數位資源類型予以數位保存，包括地理資料、網站、電視、社會科學資料集、電子期刊以及歷史資料六種數位資源類型。

「地理資訊資料」係由各州或當地政府機構產生，廣泛地被用於公共與私人部門，如：醫療保險計畫(Medicare)與醫療補助計畫(Medicaid)、稅收估價(Tax Assessment)、企業規劃(Business Planning)、家園安全(Homeland Security)與環境議題。現在有許多地圖是屬於原生數位的資料，也蘊含非常重要而豐富的內容，數位地圖資料也能夠很容易地把舊的資料覆蓋而成為新的資料，屬於具有高風險性的數位內容而有必要加以保存。有關於地理資訊資料的保存可以參考 NDIIPP 計畫伙伴：北卡羅萊納州立大學(North Carolina State University)以及加州大學聖塔芭芭拉分校(University of California at Santa Barbara)。

「網站」呈現愈來愈多政治與社會事件的重要紀錄，與日常生活息息相關。在網站資料的保存部分，NDIIPP 持續地支援數個著重於保留具有高價值、高風險的網站內容計畫。有關於網站資料的保存包括：加州數位圖書館(California Digital Library)，伊利諾大學香檳校區(University of Illinois at Urbana-Champaign)與 MINERVA 計畫。(註14)

「電視」具有廣播傳遞的本質，也是數位內容中最具有風險的一種格式，目前 NDIIPP 對於美國與外國某些國家的非商業性質節目特別感興趣，每年 8,000 個小時的節目，有 30 個國家外語新聞節目（如：埃及、伊朗、俄羅斯、南非）及公共電視節目（包括從前線(Frontline)、宗教與倫理和大自然中挑選的內容）。有關於數位電視節目的保存可以參考 NDIIPP 計畫伙伴：大眾廣播集團(Public Broadcasting System)和 SCOLA。(註15)

「社會科學資料集」蘊含大量有關於美國社會與經濟的資料，可以幫助瞭解未來政治與經濟等大環境的發展，包括宗教與國家民意資料、國際民意與社會經濟調查等，然而，目前尚有大量的社會科學資料未適當地被保存。有關於社會科學資料集的保存可以參考 NDIIPP 計畫伙伴「全球政治與社會科學聯盟出版品與會議資料」(Inter-university Consortium for

Political and Social Research，簡稱 ICPSR)，該聯盟於 1962 年成立，主要在於提供全球政治與社會科學聯盟的出版品與會議資料。(註16)

「電子期刊」是每年教育機構大量產出的一種資訊資源，其提供許多優點，例如全文檢索。但是沒有人能夠保證電子期刊的內容到許久以後的未來都還能維持取用，因為保存核心電子學術文獻所需的基礎建設規模與複雜，遠超過任何單一圖書館或機構組織的預算所能負擔，所以需要產生新的合作方式共同分擔長期保存與取用電子期刊的成本。更多有關於電子期刊的保存可以參考 Portico 和 LOCKSS(Lots of Copies keep Stuff Safe)。(註17)

「歷史資料」的產生是個人的寫作、企業的紀錄、紀錄性的回顧、新聞和許多經過多年還存在的文件。數位時代的歷史乃從現在什麼能夠被保留下來而開始，不幸地，有許多網站早期的歷史紀錄已永久遺失。(註18)

除網站類原生數位資料之外，其餘五種數位資源類型的數位保存工作，係由美國國會圖書館主導，支援其他資訊保存相關的機構執行。唯網站擷取與典藏部分，國會圖書館自行發展 MINERVA 計畫，此外亦有與加州大學數位圖書館合作「危機中的網頁：一個分散式保存國家政治文化遺產的方法」計畫，與伊利諾大學香檳校區合作開發 ECHO REPOSITORY，以及與 IA、IIPC 分別合作網站擷取與典藏等計畫。研究者以圖 5-3 說明美國國會圖書館與 NDIIPP 計畫及 MINERVA 計畫間之關連。

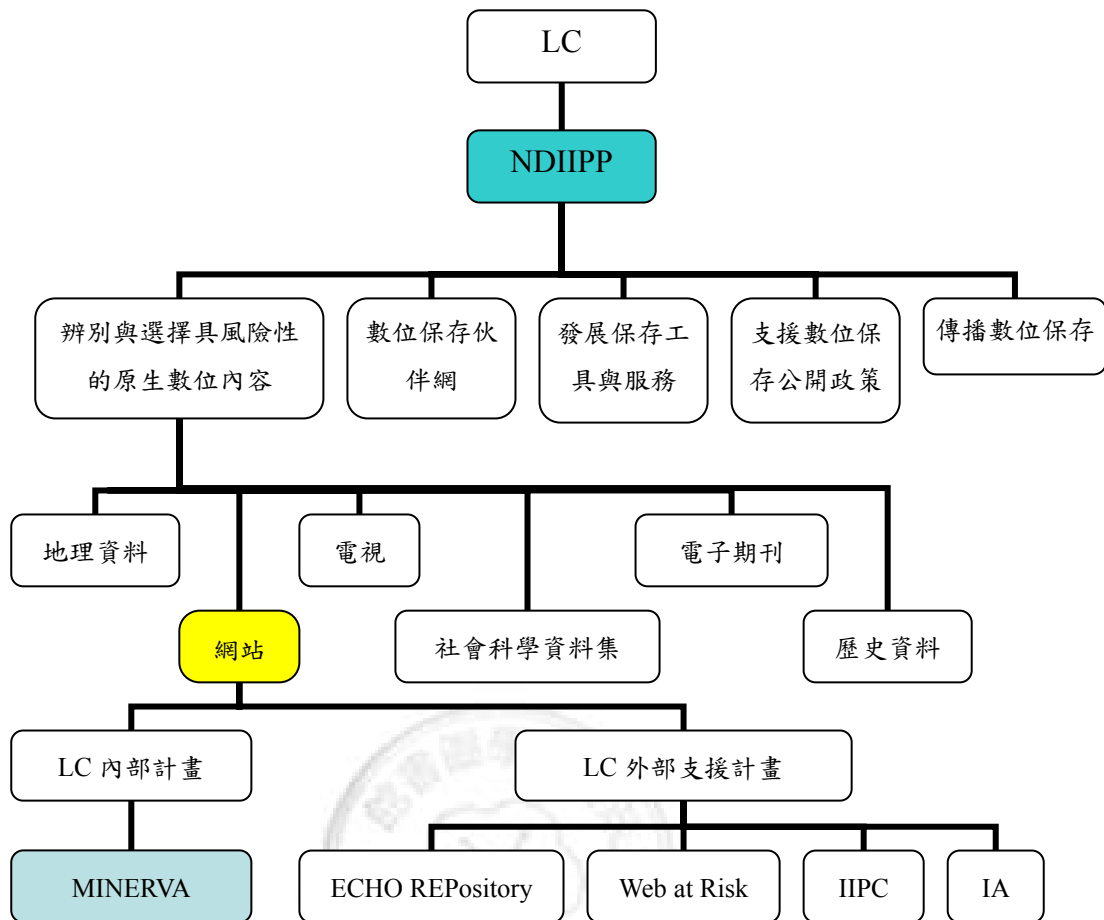


圖 5-3 美國國會圖書館數位保存與計畫關係圖

※資料來源：研究者整理

由上述可知，以美國國會圖書館為主導的數位保存計畫為 NDIIPP 計畫，而以美國國會圖書館所主導的網站典藏與保存計畫則以 MINERVA 計畫為主。

### (一) 國家數位資訊基礎建設與保存計畫(NDIIPP)

2000 年，美國國會意識到為下一代保存數位內容的重要性，又鑑於無法獨立完成所有數位保存的工作，乃於 12 月通過特別立法（公開法案 106-554）撥款一億元委託國會圖書館領導其他聯邦機構（如檔案館、博物館及其他聯邦機構等），合作發展「國家數位資訊架構和保存計畫」。該計畫主要目的在於發展一個國家策略、蒐集、典藏與保存大量產生的數位內容，特別是針對只有產生數位格式而無實體之資料，以提供現在與未來世代的社會大眾、使用者、亦或是研究者來取用。

國會圖書館透過 NDIIPP 計畫，與來自於大專院校、圖書館、檔案館、聯邦機構和商業內容與科技組織等伙伴合作，以達成下列五項策略性目標：

1. **辨識與選擇具風險性的原生數位內容**，如僅以數位形式產生的地圖、電影、音樂記錄、網站和資料庫。
2. **建立與支援數位保存伙伴網**。由於保存數位文化遺產的任務太過龐大，並非單一圖書館或者檔案館能夠獨立勝任的工作，只有藉著共同合作的機會，才能為現在與未來的使用者保留更多、更豐富的知識內容。目前，NDIIPP 計畫的各個伙伴各有不同領域、不同目的，有負責蒐集與保存內容為主者，有以數位保存研究為主者，有以發展保存數位內容所需的工具與服務為主者。此外，美國國會圖書館亦召集會議，成員為來自於美國 50 州、哥倫比亞行政區及 3 個美國領地的代表，探討各州當地數位文件的保存問題。最後，美國國會圖書館身為伙伴網之一員，亦積極參與國內與國際其他組織，以達成數位保存計畫成功的里程碑。(註19)
3. **發展與應用在數位保存用途之工具、模型、方法和服務**。科技的轉變雖然十分快速，但是對於數位資訊的保存亦十分有用，發展有關數位保存活動的工具與服務，將可幫助圖書館與檔案館管理、儲存與檢索數位資訊。
4. **鼓勵並支援數位保存公開政策**。如何保存並使數位資料能夠取用，是圖書館與檔案館所面臨的最大困難之一，而現在正值挑戰著作權、智慧財產權最艱困的時代，如何才能取得圖書館等保存機構與出版者、創作者間的平衡，鼓勵其支援數位保存公開的政策，將是必然面對的問題之一。
5. **傳播數位保存重要的觀念**。資訊科技時代，數位相機、網際網路與攜帶式音樂裝置等，使每一個人都有創作與貢獻數位資訊的機會和能力，本計畫致力於宣傳保存自己的原生數位資訊是每個人的工作與責任的觀念。(註20)

NDIIPP 計畫最初的重點著重在建立國家網絡，透過該網絡共同發展及測試數位保存架構，並支援進階的數位保存研究。此外，與所有和 NDIIPP 計畫活動相關的利害關係者(stakeholders)互通聯繫也是計畫開始之要項之一。NDIIPP 計畫期望透過與伙伴間的合作關係進行數位保存的活動與計畫，因此，國會圖書館乃與幾個重要聯邦機構達成協議，共同建立伙伴關係，包括：國家檔案和文件管理局（National Archives and Records Administration，簡稱 NARA）、國家醫學圖書館、國家農業圖書館，此外亦與資訊科技公司、大學、非營利實體（如：RLG、OCLC、CLIR）等機構合作。

除上述五項目標之外，NDIIPP 未來也將持續探究數位內容的選擇、館藏策略、與保存相關的智慧財產權議題，與數位保存活動相關之經濟持續性議題，及數位保存的技術架構等。（註21）

2002 年，國會贊助「保存我們的數位遺產：國家數位資訊基礎建設與保存計畫方案」(Preserving Our Digital Heritage: Plan for the National Digital Information Infrastructure and Preservation Program)，該計畫的方針原則與目標構成 NDIIPP 計畫，並致使邁向成功的里程碑，較為重要的事件以及近幾年來的發現包括：(1)將有遺失風險的數位內容進行辨別與蒐集；(2)召集國際共同保護數位資訊；(3)發展能夠妥善照顧數位資訊的新工具與程序。其近幾年來的重要發展整理如下表 5-2「Preserving Our Digital Heritage」計畫成果。（註22）

表 5-2 美國國會圖書館 Preserving Our Digital Heritage 計畫成果

重要發展目標	成果說明	
	年代	事件
1. 辨別與蒐集具有遺失風險的數位內容	2004 年	NDIIPP 成員從網站、數位電視、商業性檔案 (Business Records)、資料庫、geospatial data、文化遺產文件與電子出版品等具有遺失風險的資源進行蒐集與選擇，截至 2006 年已有超過 66TB(Terabytes)的數位檔案被鑑定、選擇與保存。
	2005 年	Portico，一個非營利電子典藏服務(Electronic Archiving Service)提供以電子形式出版的學術資源典藏服務。
2. 召集國際保護數位資	年代	事件

訊	2004 年	美國國會圖書館建立一個正式的 NDIIPP 數位保存伙伴網，由八個數位保存聯盟所組成，每一個聯盟都各擁有許多機構會員，他們利用 NDIIPP 計畫的研究與當地符合的資源結合，蒐集與保存有風險的數位內容，例如：地理資訊、網站、數位電視、社會科學資料集、從.com 時代與文化遺產文件開始的企業紀錄等數位資源。發展數位館藏與保存策略、探討支援伙伴間運作的協定與標準，以及支援保存工具、模式與方法的發展。
	2006 年	著作權法第 108 節研究小組(Section 108 Study Group)於洛杉磯及華盛頓 D.C. 兩地分別舉辦公開圓桌會議，蒐集如何在著作權法第 108 節中為圖書館及檔案館修訂著作權例外案的意見與看法。
3. 發展數位資訊新工具與程序	年代	事件
	2004 年	<ul style="list-style-type: none"> <li>● 國家科學基金會 (National Science Foundation, 簡稱 NSF) 與 NDIIPP 合作研究數位保存事項，命名為「It's About Time」計畫。</li> <li>● 哈佛、史丹佛、老道明及約翰霍普金斯四所大學參與保存測試，模擬長期挑戰數位內容科技的成效。</li> </ul>
	2005 年	NSF 與圖書館聯合頒發 10 個研究獎金。
	2006 年	史丹佛大學成為 NDIIPP 計畫技術架構發展的合作伙伴。

除上述成果之外，另有下列三方面之成果：(1)建立數位保存伙伴；(2)技術架構；(3)數位保存研究。首先，在建立數位保存伙伴的成果發展方面，2004 年美國國會圖書館為數位保存活動建立了八個聯盟，每一個聯盟還包括多個機構會員。此八個聯盟結合 NDIIPP 計畫的投資以及當地的資源，其將進行辨識與保存有風險的數位內容，發展大規模的館藏與保存策略、探討支援伙伴間運作所需的協定與標準、並支援保存工具、模型和方法的發展。每個子計畫也都將探討內容選擇與館藏策略，探測與保存相關的智慧財產權議題，關注經濟是否足以支援數位保存活動所需，以及技術架構之探討等。所有計畫的參與者都是採合作性互動的方式，共享發展與其經

驗。有關於 NDIIPP 計畫的八個聯盟與其執行的計畫，整理如表 5-3 NDIIPP 計畫數位保存伙伴。(註23)

表 5-3 NDIIPP 計畫數位保存伙伴

	主要負責機構	成員(Partners)	計畫主旨
1.	加州大學加州數位圖書館(California Digital Library)	紐約大學；北德州大學（圖書館及數位知識中心） 合作者：聖地牙哥高速電算中心（San Diego Supercomputer Center，簡稱 SDSC）；史丹佛大學資訊科學系與美國昇陽電腦公司(Sun Microsystems Inc.)	發展網站典藏工具的計畫，以使圖書館蒐集政府與政治的網路資訊資源，其欲建立的館藏將著重於當地政治活動與發展，例如：2003 年加州州長罷免選舉。
2.	加州大學聖塔芭芭拉分校	史丹佛大學	建置國家地理資訊聯邦數位典藏庫(National Geospatial Federated Digital Repository)，設計一個基礎架構與蒐集合跨不同地理資訊形式的資料。該典藏庫將保存對於歷史研究、科學研究、環境政策、城市人口研究、人口普查建設與分析、及其他需要美國地理資訊的領域
3.	教育廣播法人 (Educational Broadcasting Corporation，簡稱 EBC)	WGBH 教育基金會；公共廣播服務 (Public Broadcasting Service，簡稱 PBS)；紐約大學。	建立保存數位公共電視節目的程序、結構與國家標準。EBC 和 WGBH 是美國公共電視內容最大的生產者，藉由 PBS 的服務使其節目可以放送到美國沿岸各地。
4.	美國艾默雷大學 (Emory University)	路易斯威爾大學圖書館；維吉尼亞理工學院與其州立大學圖書館；佛羅里達州立大學；Auburn 大學圖書館；喬治亞學院科技圖書館與資訊中心。	此合作計畫主要在建置美國南方文化遺產的 MetaArchive( <a href="http://www.metaarchive.org">http://www.metaarchive.org</a> )，建置分散式數位保存網，將有關美國南方的文化與歷史中重要且有風險的內容加以典藏與保存。
5.	伊利諾大學香檳分校圖書館、伊利諾大學圖書資訊研究所、國家高速應用程式中心	OCLC； <a href="#">法國杜夫特大學(Tufts University)</a> <a href="#">柏修斯數位圖書館計畫(Perseus Project)</a> ；密西根州立大學圖書館；以及亞利	此計畫將發展決定擷取與保存何種數位資料的標準，因為不是所有數位資料都能或都應該被保存，而資料將包含聲音與影音記錄、歷史照片、參與者各州的網路政府出版品，

	(National Center for Supercomputer Applications)	桑納、康乃狄克、伊利諾、北卡羅萊納與威斯康辛州立大學圖書館聯盟。	以及主要和次要的歷史資料。
6.	馬里蘭大學史密斯商學院(Robert H. Smith School of Business)	喬治梅森大學歷史與新媒體中心；Gallivan, Gallivan and O'Melia LLC；Snyder, Miller, Orton Lawyers LLP；以及 IA。	此計畫將保存與美國.com 時代起的網際網路貿易行為相關之數位資料，尤以 1994 至 2001 年有關於美國商業文化的數位內容為主。並提供下列網址為取用管道： <a href="http://www.businessplanarchive.org">http://www.businessplanarchive.org</a> <a href="http://www.dotcomarchive.org">http://www.dotcomarchive.org</a>
7.	密西根大學全球政治與社會科學聯盟出版品與會議資料 (Inter-university Consortium for Political and Social Research)	康乃狄克大學公共意見研究 Roper 中心；the Howard W. Odum Institute for Research in Social Science at the University of North Carolina-Chapel Hill；the Henry A. Murray Research Center at the Radcliffe Institute of Harvard；NARA 電子與特殊資料文件服務部；Harvard-MIT 資料中心。	辨識、徵集與保存與社會科學相關的數位資源，例如投票記錄、大規模的家族成長與收入調查、工廠倒閉事件的研究、老年人照護的需求研究等。
8.	北卡羅萊納州立大學圖書館	北卡羅萊納地理資訊與分析中心	此計畫將蒐集與保存數位地理資料資源，包括從各州與北卡羅萊納州當地政府機構蒐集而來的數位化地圖。

NDIIPP 計畫的第二項成果是在技術架構方面，在剛開始規劃的階段，NDIIPP 從私人與公共部門召集科技人才組成四次工作會議，在工作會議之後產生保存技術結構架構，欲作為指引國家分散式網絡的發展。該技術結構的原則是必須支援機構與機構間之關係，區分保存與取用的功能以保護作者智慧財產權，且該架構必須經常修正、時常進行部分的更新，使用廣泛採用的標準與協定。目前技術架構已發展至第二版，可於 <http://www.digitalpreservation.gov> 網站取用。（註24）

NDIIPP 計畫的第三項成果是在數位保存研究方面，美國國會圖書館與國家科學基金會（National Science Foundation，簡稱 NSF）合作建立數



位典藏與長期保存 DIGARCH 研究計畫，獲得 3 百萬美元執行該項計畫之先導研究。該計畫將發展著重下列主題：(1)從數以百計的深海潛航任務中蒐集與保存豐富的海洋資料；(2)使描述數位物件能夠自動化的方法，以及將數位物件置放於安全的典藏儲存環境的方法；(3)測試如何在數位影音在一開始建立時即可保存的方法；(4)保存複雜的三度空間數位內容。(註25)

## (二) 網際網路電子資源虛擬典藏計畫(MINERVA)

國會圖書館向來以保存國家文化作品並提供取用為己任，圖書館傳統的徵集、編目、保存與服務館藏資料等功能，促進了教育、學術、與研究各方面的發展，現在更要延伸至數位資料、網站之蒐集、編目、保存與服務等功能。(註26) 乃由於世界文化與智慧不斷地以數位化的格式產生，且有些數位資源並無其他實體格式的存在，尤其是網際網路資源，萬一遺失數位內容或者一旦連結失效，資料內容可能再也找不回來，無疑是每一個使用者的損失。美國研究顯示網站資源穩定地成長中，但同時也持續地在消失中，網頁的平均壽命僅有 44 天，並不長久，其根據統計發現如果在 1998 年能夠找得到的網頁，隔年即找不到該網頁的情形約佔 44%。(註 27) 故國會圖書館為使其資源能夠被國會會員與美國人民長久取用、可及，以支援全球知識館藏，培養未來世代子孫的創造力與學習，乃於 2000 年正式發展「網際網路電子資源虛擬典藏計畫」(MINERVA)。計畫係國會圖書館一個網站典藏先導計畫，建立蒐集與保存重要的網際網路資料。國會圖書館組成一個包含編目員、法律諮詢專業人員、資訊科技提供者、技術專家、參考服務館員、策略計畫發展辦公室等的多學科專家小組，專門研究評估、選擇、蒐集、編目、提供取用、保存網際網路資源的方法，作為未來研究之參考。(註28) 計畫其他合作伙伴尚有：(註29) (註30)

- (1) 國際網際網路保存聯盟(IIPC)：負責蒐集與保存全世界豐富的網際網路內容。
- (2) 加州大學數位圖書館：負責「危機中的網頁：一個分散式保存國家政治文化遺產的方法」(The Web at Risk: A Distributed Approach to Preserving our Nation's Political Cultural Heritage) 計畫。
- (3) 伊利諾大學香檳校區，OCLC：The ECHO REpositoy。負責典藏品選擇方法與工具之發展。

- (4) 網際網路典藏(Internet Archive)：負責網際網路典藏工具的發展；以及測試已蒐集的網站內容之儲存、資料維護與取用。
- (5) WebArchivist.org：是一研究型組織，包括華盛頓大學通訊與市民公約中心 (University of Washington, Center for Communications and Civic Engagement, 簡稱 UW, CCCF)，以及紐約州立大學科技學會 (State University of New York, Institute of Technology, 簡稱 SUNY-IT)。目的在於發展辨識、蒐集、編目與分析大規模網站物件的典藏系統。負責「2002 年美國選舉」與「911 恐怖攻擊事件」兩個主題子計畫網站典藏之編目。

MINERVA 計畫的發展程序包括：(1)規劃館藏：須先定義「事件」，以事件為主題進行網站典藏計畫、確定經費是否足以支援、發展工作次序與流程（包括訂定館藏範圍、館藏大小、蒐集開始與結束的時間、定義網站 URLs 的種類以便於選擇、擷取頻率與深廣度）；(2)檢視與網站典藏相關的法律問題；(3)選擇網站：依事件不同而發展其適合的館藏政策說明、找到符合選擇標準且特別的網站、找出網站資訊聯絡者、擁有者或產生者以協商權利管理 (Rights Management) 問題；(4)註記與權限 (Notifications/Permission)；(5)技術分析：決定最好的款目點 (Point of Entry)、爬梳品質的保證等；(6)利用擷取工具爬梳；(7)將擷取並選為典藏的網站內容利用 MODS 描述性後設資料進行編目，記錄於美國國會圖書館整合型圖書館系統 (Integrated Library System, 簡稱 ILS) 中，使編目記錄得以呈現於國會圖書館 OPAC 中；(8)發展取用介面；及(9)取用，其發展流程如圖 5-4 MINERVA 發展過程圖所示。

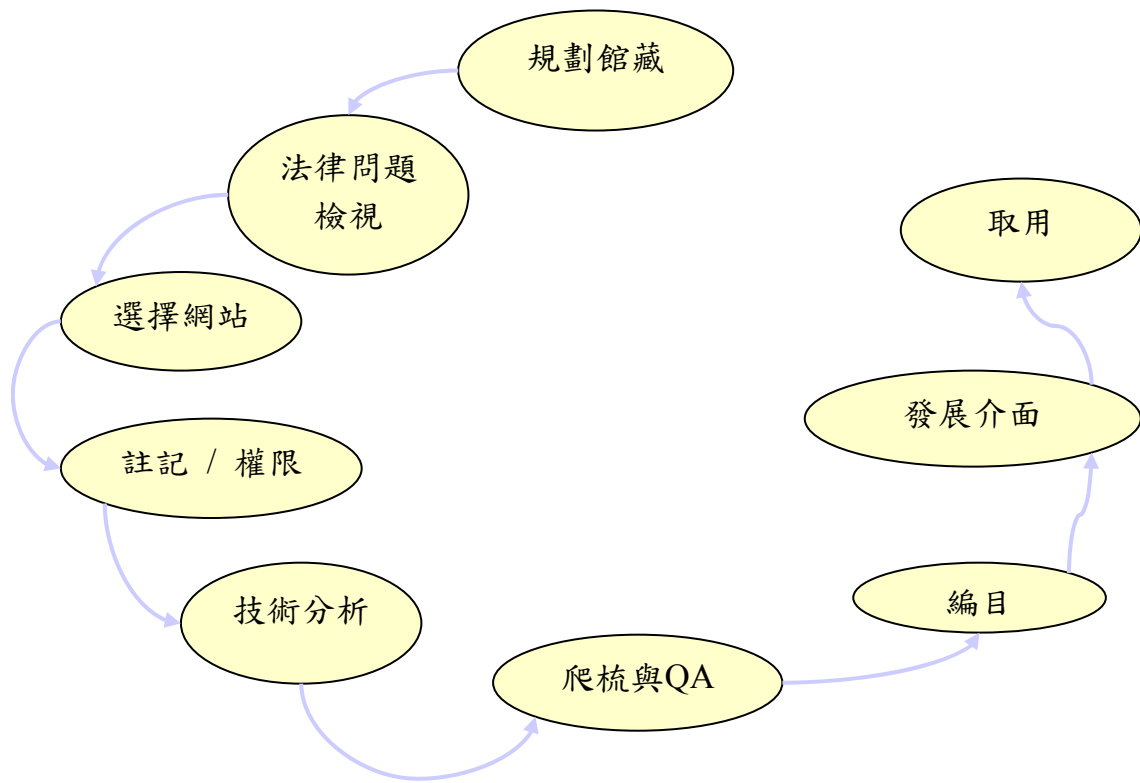


圖 5-4 MINERVA 發展過程圖

全世界的國家圖書館利用許多種不同的方法去選擇與典藏各自國家的網站出版品，其蒐集的方式與種類有：選擇性的蒐集、週期性蒐集該國所有網域之數位出版品、主題式蒐集(Thematic Collecting)、以及寄存館藏建立四種。透過網站擷取工具 HTTrack，美國國會圖書館則是依照特定主題建立其網站的典藏，而自計畫開始至今已擷取超過 37,578 個網站，網站主題包括 2000、2002 及 2004 年的美國國家及總統大舉、伊拉克戰爭以及 911 恐怖攻擊事件等。研究者將 2000 年 MINERVA 計畫開始迄今所蒐集的網站主題、蒐集日期與內容整理如表 5-4 MINERVA 網站典藏主題計畫一覽表。(註31)

表 5-4 MINERVA 網站典藏主題計畫一覽表

網站主題	蒐集日期	蒐集內容
2000 年選舉網站典藏	2000 年 8 月至 2001 年 1 月	美國國會、州長與總統選舉；政黨；政府；提倡；大眾意見；以及各種與 2000 年美國選舉相關的網站資源。
911 恐怖攻擊事件網站典藏	2001 年 9-12 月	個人與團體的網站、美國新聞報刊(press)及機構、2001 年 9 月 11 日恐怖組織攻擊事件發生後全球相關的報導。
2002 年冬季奧運	2002 年 2 月	與 2002 年在美國猶他州鹽湖城所舉行的奧林

		匹克運動會相關之網站。
2002 年選舉網站典藏	2002 年 7-11 月	國會、州長與總統大選；政黨；政府；提倡；部落格；大眾見以及各種與 2002 年美國選舉相關的網站資源。
911 紀念	2002 年 9 月	911 攻擊事件週年紀念。
第 107 屆國會	2002 年 12 月 12 日	國會與立法委員之官方網站。
2003 年伊拉克戰爭	2003 年 3 月迄今	與伊拉克戰爭有關的公開政策網站。
2004 年選舉	2003 年 10 月至 2004 年 12 月	國會、州長與總統大選；政黨；政府；提倡；部落格；大眾見以及各種與 2004 年美國選舉相關的網站資源。
美國國會的網站	2003 年 12 月迄今	白宮、參議院、國會委員會以及與國會相關的網站。
羅馬教宗的變遷	2005 年 4-5 月	約翰保路二世逝世的回應，以及新教宗的選舉。
最高法院 Supreme Court	2005 年 8 月至 2006 年 2 月	最高法院提名與任命流程。
卡翠那颶風 (Hurricane Katrina)	2005 年 9-10 月	與 IA(Internet Archive)、加州數位圖書館整理的天災卡翠那颶風相關訊息、實況報導及回應。
蘇丹達夫(Darfur)地區的危機	2006 年 2-9 月	政府、國際組織和美國一般大眾至全球對於蘇丹達夫地區危機的組織、新聞報導與回應。
一般館藏典藏先導計畫	2006 年 4-9 月	非主題性館藏網站典藏先導計畫。
機構組織網站的手稿典藏	2006 年 4-9 月	包括公民權利(Civil Rights)、政治提倡團體、專業與名譽組織、紀念性團體語研究教育組織所貢獻的網站。
紙本與照片徵集	2006 年 4-9 月	照片、圖表與其他視覺資料網站(可加強 P&P 館藏者)
2006 年總統大選	2006 年 5-11 月	國會、州長與總統大選；政黨；政府；提倡；部落格；大眾見以及各種與 2006 年美國選舉相關的網站資源。

最後，研究者整理與國會圖書館數位保存計畫相關的重要紀事與發展如表 5-5 美國國會圖書館數位保存大事紀。

表 5-5 美國國會圖書館數位保存大事紀

時間	事件
1909 年	凡受到著作權法保護的作者，在作品出版以後的 3 個月內，皆有義務寄存兩本作品之複本交與國會圖書館典藏。
1976 年	1976 年著作權法
1988 年	開始寄存磁帶、光碟和微縮資料
1990 年	美國國會圖書館發展「美國記憶」數位典藏計畫
1993 年	建立 CD-ROMs 的自願寄存協議
1994 年	美國國會圖書館發展「國家數位圖書館計畫」(NDLP)
2000 年	2000 年千禧著作權法
2000 年 8 月至 2001 年 1 月	MINERVA 子計畫 - 2000 年選舉網站典藏
2000 年 12 月	美國國會圖書館發展「國家數位資訊基礎建設與保存計畫」(NDIIPP)
2001 年	第一本電子書透過電子著作權法寄存
	美國國會圖書館成立策略計畫辦公室(OSI)
2001 年 9-12 月	MINERVA 子計畫 - 911 恐怖攻擊事件網站典藏
2002 年	美國國會圖書館發展「保存我們的數位遺產：國家數位資訊基礎建設與保存計畫方案」
2002 年 2 月	MINERVA 子計畫 - 2002 年冬季奧運
2002 年 7-11 月	MINERVA 子計畫 - 2002 年選舉網站典藏
2002 年 9 月	MINERVA 子計畫 - 911 週年紀念
2002 年 12 月 12 日	MINERVA 子計畫 - 第 107 屆國會
2003 年 3 月迄今	MINERVA 子計畫 - 2003 年伊拉克戰爭
2003 年 4 月與 5 月初	美國國會圖書館與全球企業網路分別在加州柏克萊與紐約舉辦兩場有關 NDIIPP 架構的研討會，其後再於 5 月 30 日及 7 月 28、29 日在紐約召開小型焦點座談，最後產生 NDIIPP Architecture Version 0.2 報告書。
2003 年 6 月	OCLC 與 RLG 組成 PREMIS 工作小組
2003 年 7 月	美國國會圖書館與澳洲、加拿大、丹麥、芬蘭、法國、冰島、義大利、挪威、瑞典、英國等國家圖書館，及美國網際網路典藏(IA)成立國際網際網路保存聯盟(IIPC)。
2003 年 10 月至 2004 年 12 月	MINERVA 子計畫 - 2004 年選舉
2003 年 12 月迄今	MINERVA 子計畫 - 美國國會的網站

2004 年	美國國會圖書館策略計畫辦公室組成網站擷取小組，支援國家數位資訊基礎建設與保存計畫的策略性目標，以管理和支援急待典藏之數位內容。(註32)
	NDIIPP 計畫成員開始蒐集六種原生數位資料
	NDIIPP 計畫與國家科學基金會(NSF)合作研究數位保存工具與程序，發展「It's about Time」計畫。
	哈佛、史丹佛、老道明與約翰霍普金斯參與保存工具測試，模擬長期挑戰數位內容科技的成效。
2005 年	非營利電子典藏服務 Portico 開始提供電子版學術資源典藏服務
	「選擇與管理網站內容擷取」(Selecting and Managing Content Captured from the Web)先導計畫。(註33)
2005 年 4-5 月	MINERVA 子計畫 - 羅馬教宗的變遷
2005 年 5 月 4 日	美國國會圖書館的 NDIIPP 計畫與國家科學基金會(NSF)提供 10 所美國大學共 300 萬美金進行數位保存初探性研究。
2005 年 8 月	圖書館館員著手進行一年期的計畫，研究與決定選擇網站資源的選擇指南。
2005 年 8 月至 2006 年 2 月	MINERVA 子計畫 - 最高法院
2005 年 9-10 月	MINERVA 子計畫 - 卡翠那颶風
2006 年	美國國會圖書館著作權法第 108 節研究小組成立，並於洛杉磯、華盛頓兩地分別舉辦公開圓桌會議，蒐集著作權例外案的意見與看法。
2006 年 2-9 月	MINERVA 子計畫 - 蘇丹達夫地區的危機
2006 年 4-9 月	MINERVA 子計畫 - 紙本與照片徵集
	MINERVA 子計畫 - 機構組織網站的手稿典藏
	MINERVA 子計畫 - 一般館藏典藏先導計畫
2006 年 5-11 月	MINERVA 子計畫 - 2006 年總統大選

### 三、組織

1998 年，受到資訊科技變遷的影響，國會圖書館計畫將機構改組。2000 年，國家研究協會(National Research Council)的 Computer Science and Telecommunications Board(CSTB)提出有關於美國國會圖書館資訊科技的未來之研究與策略性的建議，內容尤其關心數位與類比館藏的整合，以及如何與大量的數位圖書館科技、系統和計畫互動交流等議題。2001 年，受

到該調查報告的影響，乃成立策略計畫辦公室(OSI)。(註34)

美國國會圖書館自從 1994 年執行國家數位圖書館計畫 (National Digital Library Program, 簡稱 NDLP) 以來，最主要的其中一項成果即為美國記憶計畫，該計畫是一個大型數位化計畫，使美國國會圖書館的館藏可以透過網際網路提供全球的使用者取用。NDLP 計畫中所建立的數位化過程、永久辨識號(Persistent Identification)與儲存管理(Storage Management)之實務廣泛地被應用到其他有關數位內容取用的計畫，例如：全球入口開道(Global Gateway)、音樂領域的數位館藏、印刷品與照片的線上編目、退役軍人歷史計畫(Veterans History Project)。(註35)

2000 年 12 月國會撥款 1 億美元，直指美國國會圖書館負責發展 NDIIPP 計畫，要求美國國會圖書館與其他聯邦或非聯邦實體機構合作，找出有責任蒐集、維護與提供取用數位資料的圖書館與其他組織機構，共同形成一個國家網(National Network)。一旦此策略發展，將可開始著重於發展長期保存數位資源的政策、程序與策略（當然亦包括美國國會圖書館所需的技術基礎建設等）。而策略計畫辦公室(OSI)即作為負責管理 NDIIPP 計畫的服務單位。(註36)此外，策略計畫辦公室亦負責所有與圖書館及國家數位文化資產長期保存計畫相關的策略計畫，其下有資訊科技服務組 (ITS)，負責美國國會圖書館電腦與電信通訊基礎建設，並支援服務單位及其館員的科技需求。

美國國會圖書館的所有單位組織在數位保存活動皆有一定相關的業務部分，OSI 為了使各單位機構能夠在 NDIIPP 等計畫之數位保存工作實務上，科技整合並彼此合作。乃於 2002 年，召集各部門主管組成「數位執行管理小組」(Digital Executive Oversight Group, 簡稱 DEOG) 定期舉行會議，研究與資源檢查、判斷和分配有關的數位計畫以及資訊科技策略，並支援國會圖書館內部的數位策略活動，該小組成員共有 6 位，分別為：策略計畫辦公室聯絡館員 Laura Campbell、館員主管 Jo Ann Jenkins、圖書館服務組聯絡館員 Deanna Marcum、法律圖書館館員 Rubens Medina、國會研究服務部主管 Daniel Mulhollan、著作權註冊負責人 Marybeth Peters。(註37) 目前，OSI 和資訊科技服務組總共各有 96 與 203 位館員，其他各部門與數位保存活動相關的館員則視情況與需求而分派。在 MINERVA 計畫的團隊組織方面，則有團隊主管 Diane Kresh 所帶領的各學科領域所組成的專家小組，包含專業編目人員、法律諮詢專業人員、資訊

科技提供者、技術專家、參考服務館員、策略計畫發展辦公室人員等參與。  
(註38)

### 第三節 數位典藏

根據本論文研究範圍與限制，在數位典藏與取用技術的部分，以網站典藏計畫為主要的觀察對象。故本節將以美國國會圖書館之網站典藏計畫「網際網路電子資源虛擬典藏計畫」(Mapping the INternet Electronic Resources Virtual Archive，簡稱 MINERVA) 為主要的研究與觀察對象。

#### 一、數位資源蒐集與選擇

##### (一) 數位館藏類型與範圍

「國家數位資訊基礎建設與保存計畫」的主要目標，在於蒐集與保存具有高度價值、可能正處於遺失或損壞風險中、並且對於國家與國會將是有用的數位內容。所謂「高風險性」(High-Risk)的內容，係指現在還未被保留起來，且可能會永遠遺失或損壞者，它們不同於實體形式的資料，可能在 20 年、數世紀以後還能維持目前的現況，而數位格式中的內容卻很容易被改變或甚至遺失。因為並非所有產生於數位形式的資訊都能被保留下來，因此 NDIIPP 計畫也僅著重於蒐集與保留某些對於今日或未來世代有用的特定類型之數位資訊，如：地理資訊資料(Geospatial Data)、網站、數位電視、社會科學資料集(Social science datasets)、電子期刊(E-Journals)、歷史資料(Historical materials)等。(註39)

在網站典藏的部分，美國國會圖書館的 MINERVA 計畫依訂定事件為主題以蒐集網站資源，依循國會圖書館館藏發展與政策作為各子計畫蒐集網站時的標準與參考。在館藏蒐集範圍方面，乃因國會圖書館蒐集任何格式之資料以支援全球性的館藏內容為目的，故網站資源的擷取與典藏也不計較資料的格式或類型。一般而言，所蒐集的數位資料將包括網站、首頁、或網頁上的音樂視訊資料、照片、地圖、以及包含在該主題之內的相關項目皆可蒐集。

##### (二) 網站蒐集與選擇標準



美國國會圖書館在 2000 年所進行的 MINERVA 計畫，其網路資源選擇與蒐集係參考下列四份館藏選擇標準之文件訂定各主題計畫的選擇標準：

1. 美國國會圖書館電子資源館藏發展說明，網址如下：  
<http://lcweb.loc.gov/acq/devpol/electron.html>。1999 年美國國會圖書館電子資源館藏發展中提到館藏蒐集範圍包含網頁、網站、電子資源、以及聲音視訊資料、照片、印刷品、地圖等以電子出版之資料。圖書館選擇電子資源作為永久館藏，而在選擇電子出版品時應考慮作品之內容、附加價值、是否易於使用、維護、標準、設備與輸出等問題。(註40) 2004 年更新 1999 年電子資源館藏政策說明的內容，提到圖書館徵集電子資源要考慮資訊的有用性、資訊提供者的信譽聲望、資訊的獨特性、學術內容、內容是否僅以數位形式呈現、是否為易逝的資源(如灰色文獻或機密文獻)；此外，亦要考慮資料的內容、取用性、附加價值、標準、永久保管(Permanent Retention)、使用者介面與導航性。(註41)；
2. 網站擷取與典藏(Web Site Capture and Archiving)，網址  
<http://lcweb.loc.gov/acq/devpol/webarchive.html>，此指南僅提供館內館員以上人員取用，有權限制而無法閱讀。(註42)
3. 網站擷取與典藏特別指南 (Web Site Capture and Archiving Specific Guidelines)，網址  
<http://www.loc.gov/minerva/internal/info.html>。此指南僅提供館內館員以上人員取用，有權限制而無法閱讀。
4. Questions for Selectors for BEOnline Plus，網址  
<http://www.loc.gov/rr/business/beonline/>。BEOnline 始於 1996 年，是美國國會圖書館編目部門書目強化諮詢組 (Bibliographic Enrichment Advisory Team, 簡稱 BEAT) 的實驗性計畫，用意在探討企業家與小型企業遠端資源取用與書目控制的方法。BEOnline Plus 的選擇標準提到其蒐集的資源包括：(1)線上出版的印刷手冊與連續性出版品 (期刊、新聞報紙與電子報)；(2)有提供經常更新的資訊資源，例如單篇出版品(loose leaf publications)、工作報告與電子報；(3)提供上述資料類型的網

站。(註43)而選擇網站的人員必須考量網站的取用性、權威性、內容、使用者介面與導航性等問題。(註44)

在選擇標準方面，MINERVA 計畫先是由 Recommending Officers 與 MINERVA 計畫小組協商並決定所要擷取的網站為何，以選擇性的方式蒐集，而非大批擷取網站，擷取頻率則是不定期地進行擷取，擷取內容則不限何種格式之資源，選擇標準主要為：(1)資料對於現在或未來國會與研究者的資訊需求方面應是有幫助的資料；(2)資料具有獨特性；(3)富有學術內容；(4)具有遺失風險者；(5)即時性資訊。

最後，研究者將 MINERVA 計畫中，各主題子計畫網站資源之蒐集選擇標準，以及目前網站與網頁的統計數量，整理如下表 5-6 MINERVA 子計畫網站選擇標準與數量一覽表（未列於表中則代表尚無法取得相關資料）。

表 5-6 MINERVA 子計畫網站選擇標準與數量一覽表

子計畫	選擇標準
<p><b>第 107 屆國會</b> 網站 588 個</p>	<ul style="list-style-type: none"> <li>● 範圍：國會會員與委員會的官方網站。</li> <li>● 蒐集期間：2002 年 12 月 12 日</li> <li>● 徵集深度：完整網站</li> <li>● 徵集廣度：至少一個外部連結(Links followed out to one external level)</li> <li>● 擷取頻率：一次擷取</li> <li>● Robots.txt blocked sites</li> <li>● Web sites were crawled that used robots.txt blocking</li> </ul>
<p><b>2002 年選舉網站典藏</b> 網站 4,000 個 網頁&gt;70,000,000 個</p>	<ul style="list-style-type: none"> <li>● 範圍：美國 2002 中期國會選舉、州長選舉、美國 15 個主要城市市長選舉相關網站。</li> <li>● 蒐集期間：2002 年 7 月 1 日至 11 月 30 日。</li> <li>● 徵集深度：完整網站</li> <li>● 徵集廣度：至少一個外部連結(Links followed out to one external level)</li> <li>● 擷取頻率：候選人網站每日擷取，其他網站擷取頻率如后。</li> <li>● 選舉日：每小時至少擷取約 1,800 個網頁。</li> <li>● Robots.txt blocked sites</li> </ul>

	<ul style="list-style-type: none"> <li>● Web sites were crawled that used robots.txt blocking</li> <li>● 網站種類： <ol style="list-style-type: none"> <li>1. 候選人：參議員、州長、市長等。(每日擷取)</li> <li>2. 公民：個人公民出版的網站。(擷取次數視情況而定)</li> <li>3. Civic &amp; Advocacy：有在聯邦選舉委員會與國家黨派中註冊的財。(每日至每週擷取)</li> <li>4. 政府：House.gov；Senate.gov (每月擷取)；FEC.gov. (每週擷取)；各州與領地中央政府網站 (每月擷取)；選舉佈告版 (每週擷取)。</li> <li>5. 政黨：國家層級 (所有已註冊的政黨)、州屬層級 (僅指民主黨與共和黨) (Weekly for authorized) (每月擷取政黨委員會)</li> <li>6. 報刊：國家和各州主要都市、專業報刊或政治入口開道網站的新聞報刊。(每日至每週擷取)</li> <li>7. 公共意見：Gallup.com；Harrisinteractive.com (每日至每週擷取)。</li> <li>8. 其他：圖書館從 Webarchivist.org 和 Center for Communication and Civic Engagement (視情況選擇)</li> </ol> </li> </ul>
<p><b>911 恐怖攻擊事件</b>  網站 30,000 個  網頁 331,299,192 個</p>	<ul style="list-style-type: none"> <li>● 蒐集標準：(註45) <ol style="list-style-type: none"> <li>1. 所有網站應由美國國會圖書館館員所選；</li> <li>2. 用有關 911 內容之關鍵字檢索網站 URL 字串；</li> <li>3. an additional keyword search on the first instance of the home page in the archive identified relevant content</li> </ol> </li> <li>● 蒐集期間：2001 年 9 月 11 日至 12 月 1 日。</li> <li>● 擷取頻率：每日</li> <li>● Robots.txt blocked sites</li> </ul>
<p><b>2000 年選舉網站典藏</b>  網站 800 個</p>	<ul style="list-style-type: none"> <li>● 範圍：此計畫為先導計畫。主要蒐集與典藏以美國為主與 2000 年選舉相關的網站。</li> <li>● 蒐集期間：2000 年 8 月 1 日至 2001 年 1 月 21 日。</li> <li>● 擷取頻率：每日擷取。</li> <li>● Robots.txt blocked sites</li> </ul>
<p><b>911 紀念日</b>  網站 2,313 個</p>	<p style="text-align: center;">缺</p>

<b>2002 年冬季奧運</b> 網站 70 個 網頁 18,445,502 個	缺
<b>伊拉克戰爭</b> 網站 288 個	缺
<b>2004 年選舉</b> 網站 32 個	缺
<b>網站合計約 37,578 個</b>	

※資料來源：

1. MINERVA. "About – Selection."  
 <<http://lcweb2.loc.gov/cocoon/minerva/html/107th/about-selection.html>> (Retrieved at May 17, 2007).
2. Grotke, Hayes, Guenther, Ammen. "Oh, the Places You'll Go!: Recommending in the Digital Environment." February 7, 2004.
3. MINERVA. "September 11 Web Archive Browse."  
 <<http://lcweb4.loc.gov/911/911/DrillSearch>> (Retrieved at May 17, 2007).

### (三) 網站擷取方式

在網站數位保存中，一個完整的網站視為蒐集與保存的單位，係由許多檔案所構成的集合。蒐集網站的技術是利用鏡錄程式(Mirroring Program)下載網站的快拍(Snapshot)，必須將自網站伺服器中複製的檔案典藏至電腦系統中。一個網站的快拍可能包含所有網站內的檔案，也可能只有包含文字檔，週期性地下載網站的快拍，每個快拍檔案夾帶著儲存「來源性後設資料」(Provenance Metadata)，如快拍的日期等資訊。所有網站館藏的擷取與儲存皆是採自動化的方式處理。如圖 5-5 與圖 5-6 所示。

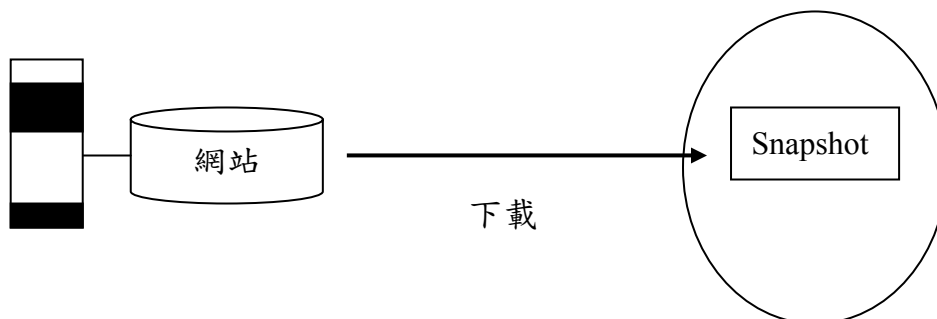


圖 5-5 快拍(Snapshot)的形成

典藏庫

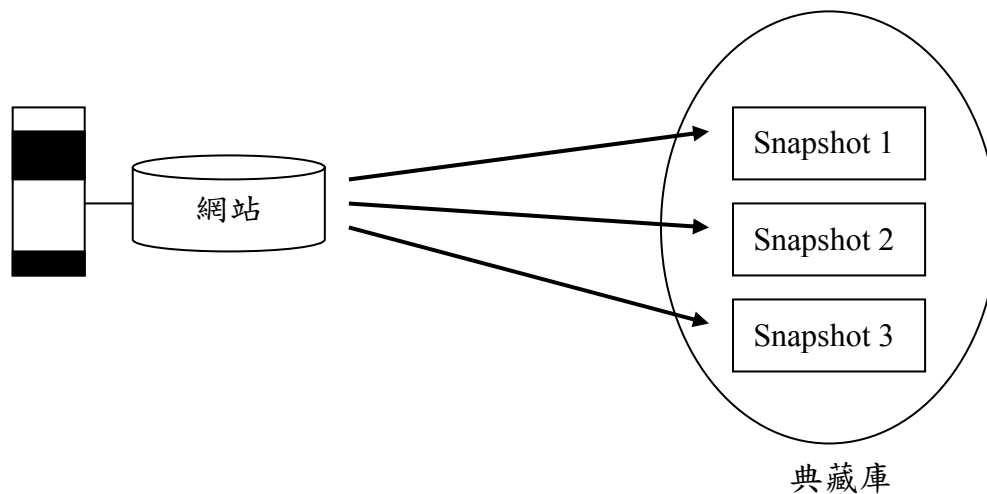


圖 5-6 週期性快拍(Snapshot)

一個快拍應該涵括所有網站部分的檔案，但有時候也需要進行選擇。網站的檔案包含了電腦執行程式與許多特殊格式的大檔案，如聲音與視訊類型檔案，如果忽略這些類型的檔案，勢必可以簡化網站蒐集的工作，但同時也失去了許多重要的資料。對於 MINERVA 計畫的雛形系統來說，過去都是使用 HTTrack 軟體來下載網站快拍檔案，使用者只要輸入網址 URL 連結，該鏡像程式可爬梳網站中所有檔案與連結，下載網站中的網頁直到整個網站都已被複製起來，其複製所有檔案包括：文字檔、影像檔、影音檔、執行程式檔、後設資料檔與樣式檔(Style Sheets)等。(註46)

現在網站蒐集主要的徵集工具主要採用網際網路典藏(IA)的網站爬梳工具「Heritrix」，Heritrix 有 heratrix、heritix、heretix、或 heratix 等不同拼法，其為 IA 的開放來源(Open-Source)軟體，是一種可延伸、以網站為基礎(Web-Scale)、具有典藏品質的網站爬梳計畫。(註47)

美國國會圖書館所採用的網站擷取工具係由其與計畫相關的參與者所共同發展，目前有下列四種領域的網站擷取工具，分別是：(1)選擇、確認(Verification)、與同意(Permissions)；(2)徵集；(3)館藏儲存與維護；(4)取用，均以「開放來源軟體」(Open-Source Software)作為其基礎架構。(註48)

而除了訂定選擇與蒐集網站的標準之外，其擷取、儲存與保存網站內容的技術需要注意下列幾項：(1)在檢索所有編碼、影像、文件及其他資料時，必要的檔案亦需保存。(2)必須擷取與保存所有來自於網站伺服器的技術性後設資料（例如：HTTP 標題）與 crawler（例如：擷取的背景脈絡、日期與時間標記、以及 crawl 環境），日期資訊尤其重要，因為日期資訊可以區辯是否重複擷取過同樣的網站，以便確認網站版本。(3)儲存內容應與傳送時一樣正確，HTML 和其他號碼皆完整無缺(Intact)而未被修改過。(4)維護平台與檔案系統的獨立性，不使用檔案系統，例如命名或時間戳記(Timestamps)紀錄技術性後設資料。（註49）

## 二、編目與 Metadata

### （一）網站資源編目

早期在網站資源編目的部分，是一個接一個地選擇、編目與典藏單一個網站。現在，MINERVA 計畫則是以主題為事件為主進行網站之蒐集與典藏，其會為每一個主題網站建立館藏層次(Collection Level)的 AACR2 和 MARC 後設資料標準紀錄，並儲存於美國國會圖書館的整合型圖書館系統(ILS)中。除了傳統的編目方式之外，又以「後設資料物件描述架構」(Metadata Object Description Schema，簡稱 MODS)支援網站題名層次(Title-Level)的描述。（註50）

MODS 係 2002-2003 年，由美國國會圖書館網路發展部、MARC 標準辦公室 (Network Development and MARC Standards Office，簡稱 NDSMO) 與後設資料專家們所共同開發。MODS 可用於各種書目元素集的架構，亦可應用在圖書館與現有 MARC 格式的相互轉換，以 XML 架構語言作為文件結構，算是一種描述性的後設資料，不但兼容現有的 MARC21 紀錄資料，並可建置原始資源的書目描述紀錄。MODS 1.0 版在 2002 年公佈並試驗 6 個月，2003 年 2 月產生修正版本 MODS 2.0，同時建置 MODS 討論群(Listserv)，提供使用者參與討論，討論期間至 2003 年 8 月止。其後，又在 2003 年 9 月公佈 MODS 3.0。目前最新的版本為 2006 年 6 月 1 日所公佈的 MODS 3.2 版。

美國國會圖書館利用 MODS 產生網站「題名層次」(Title Level)的描述性後設資料，應用於館藏層次後設資料。其在資料元素的應用彈性很大，

最高層次的資料元素共有 20 個，但是在 MINERVA 計畫中只用了 11 至 12 個，並另外著錄第 21 項「擷取日期」(Date Captured)欄位，如表 5-7 MODS 與 MINERVA 計畫使用欄位對照表。

表5-7 MODS與MINERVA計畫使用欄位對照表

欄位	MODS	MINERVA
1. TitleInfo	●	●
2. Name	●	●
3. Type of Resource	●	
4. Genre	●	●
5. Origin Info	●	
6. Language	●	●
7. Physical Description	●	●
8. Abstract	●	●
9. Table of Contents	●	
10. Target Audience	●	
11. Note	●	
12. Subject	●	●
13. Classification	●	●
14. Related Item	●	●
15. Identifier	●	●
16. Location	●	
17. Access conditions	●	●
18. Part	●	
19. Extension	●	
20. Record Info	●	
21. Date Captured		●

※資料來源：Library of Congress. “Outline of Elements and Attributes in MODS Version 3.2”. <<http://www.loc.gov/standards/mods/mods-outline.html>> (retrieved May 16, 2007).

MODS與MARC具有高度的相容性，所以可以很容易與原有的圖書館自動化整合系統結合在一起，也就是說，利用MODS作為中介，可以將網站內容的資訊轉入原有的OPAC中，讓使用者可以透過OPAC查詢到這些網

站資料，MINERVA計畫對MODS作了很好的測試，並驗證了MODS與圖書館標準的高度相容性。此外，亦驗證了MODS在資料元素選擇上的彈性，這樣的彈性讓編目人員可以更有效地產生MODS紀錄。表5-8為MINERVA計畫17個子計畫中的其中三個子計畫所使用的後設資料與欄位。

表 5-8 MINERVA 子計畫後設資料與使用欄位

子計畫	Metadata	使用的欄位
第 107 屆國會網站典藏	MODS version3.0	題名；作者；摘要；擷取日期；類型；實體描述/格式；相關款目；識別號；語言；取用條件；主題與分類號。
2002 年選舉網站典藏	MODS	題名；作者；摘要；擷取日期；類型；實體描述/格式；相關款目；識別號；語言；取用條件；主題。
911 恐怖攻擊事件	MODS	題名；作者（創作者或 issue 出版者）；摘要；主題（9 月 11 日恐怖攻擊、阿富汗戰爭、生物恐怖主義）；擷取日期；類型；語言；識別號（Active sites & Archived sites）。

「第 107 屆國會網站典藏」、「2002 年選舉網站典藏」及「911 恐怖攻擊事件」三個子計畫皆使用 MODS 作為其描述後設資料的標準，在 MODS 提供的 20 項欄位中，MINERVA 子計畫使用 8-12 項的欄位來描述其網站資料。表 5-9 為 2002 年選舉網站典藏的 MODS 紀錄範例，以 FranUlmer.com 網站來說明其描述性後設資料在館藏書目紀錄中的呈現方式。

表 5-9 MODS 紀錄範例

欄位	內容
<b>Title</b>	FranUlmer.com -- Home Page
<b>Alternative Title</b>	Fran Ulmer, Democratic Party candidate for Governor, Alaska, 2002.
<b>Name</b>	Ulmer, Fran
<b>Abstract</b>	Web site promoting the candidacy of Fran Ulmer, Democratic Party candidate for Governor, Alaska, 2002. Includes candidate biography, issue position statements, campaign contact information, privacy policy and campaign news press releases. Site features enable visitors to sign up for campaign email list, volunteer, make campaign contributions and follow links to other internet locations.



<b>Date Captured</b>	July 2, 2002 - November 30, 2002 Archived Site
<b>Subjects</b>	Elections--Alaska Governors--Alaska--Election Democratic Party (AK)
<b>Language</b>	English
<b>Genre</b>	Web Site
<b>Access Condition</b>	None
<b>Active Site</b>	<a href="http://www.franulmer.com/">http://www.franulmer.com/</a>
<b>Collection Title</b>	Election 2002 Web Archive

※資料來源：MINERVA. “MODS: Record Samples - FranUlmer.com -- Home Page (MINERVA Display Version).”  
<<http://lcweb4.loc.gov/elect2002/catalog/1405.html>> (Retrieved at May 17, 2007).

## (二) Metadata 與標準

除了 MODS 之外，為了要能夠使用數位物件，美國國會圖書館支援了多種不同的 Metadata 標準發展與調度，例如：後設資料編碼與傳遞標準 (Metadata Encoding and Transmission Standards, 簡稱 METS)、後設資料權威描述標準 (Metadata Authority Description Schema, 簡稱 MADS)、以及 Metadata for Images in XML(MIX)，並採用「PREMIS 資料辭典」(Preservation Metadata: Implementation Strategies Data Dictionary)作為其保存性後設資料標準(Preservation Metadata)。(註51)

美國保存性後設資料並非由國會圖書館開發，而是利用研究圖書館協會 (Research Libraries Group, 簡稱 RLG) 所研究的保存性後設資料。1997 年 2 月 RLG 保存性後設資料工作小組成立，小組成員包括學者、專家及圖書館專業人員，主要負責定義保存數位資源主檔 (Master File) 所需的資料項目。由於當時轉換為數位方式保存的館藏愈來愈多，而且有漸取代微縮保存的趨勢，因此欲找出可和數位資源一起保存的描述性資料，以協助轉換成數位化影像的館藏得以長久保存。1997 年 5 月工作小組檢視都柏林核心集 (Dublin Core, 簡稱 DC) 與以 USMARC 為基礎的電腦檔核心書目合作編目計畫的工作成果，企圖找出保存數位資源所需、但未被列入

DC 或 USMARC 中的後設資料資訊。1997 年 6 月工作小組提出報告草案。1997 年 11 月工作小組向顧問委員會提出正式建議，並於 1998 年 5 月正式發表 16 項保存性後設資料項目。參見表 5-10 RLG 保存性後設資料項目與範例。

表 5-10 RLG 保存性後設資料項目與範例

項目名稱	說明	格式或範例
日期 (DATE)	建檔日期	yyyyddmm
記錄者 (TRANSCRIBER)	負責記錄保存後設資料的機構名稱是必備項；個人識別資料可視情況取舍。	史丹佛大學圖書館，保存與維護部門
產生者 (PRODUCER)	數位資源實體產生的機構名稱，若經過重製手續，則記錄實際負責重製的機構，而非授權機構；可視情況加上負責產生者的個人識別資料，如掃描者等。	1.備有掃描設備的研究圖書館：史丹福大學圖書館，保存與維護部門，KES。 2.執行掃描工作的包商： Luna Imaging, Inc., 1315 Innes Place, Venice, CA 90291-3617, USA
資料擷取裝置 (CAPTURE DEVICE)	數位相機或掃描器的廠牌、規格	Kronton 3012
資料擷取細節 (CAPTURE DETAILS)	掃描器：應記錄掃描軟體相關資訊，包括軟體版本、掃描器設定、伽瑪修正功能及其他相關細節。 數位相機：應記錄鏡頭規格、焦距、光源種類等資訊。	1.Pixel Craft Pro imager 8000； 2.Nikon 24mm lens; high frequency fluorescent studio camera lights, Videsence, model PI330, with Osram 55 watt 3200 degree color temperature
變動紀錄 (CHANGE HISTORY)	檔案曾有過的修正紀錄，包括修正者／修正機構的名稱，以及修正日期	1.原始數位影像母檔，透過JWC於2001年2月6日生產的YYY軟體由TIFF v.X檔轉置成 TIFF v.X+1檔。 2.輸出檔係由原始數位母檔透過JWC於1999年4月1日出品的YYY軟體所產生，像素量不變，影像較鮮明銳利。
認證機制	用以確認電子傳輸檔案是	網際網路標準總數檢查法；Roland

(VALIDATION KEY)	否正確之機制，通常是一個數字，用來確認所傳輸的檔案是否就是後設資料所描述的檔案，即所謂「檢查碼」	總數檢查法
加密 (ENCRYPTION)	資料傳送前加密方式	RSA 公鑰加密系統
浮水印 (WATERMARK)	指出是否建立數位指紋，可供辨識影像所有者，並防止檔案遭未經授權使用。	Watermark by Digimarc Professional ; Watermark by Invisible Ink for Images
解析度 (RESOLUTION)	由像素大小標出資料解析度	4096 x 6144 pixels ; 600 dpi ; 320 dpi
壓縮 (COMPRESSION)	指出檔案是否曾壓縮，若有，應標明壓縮程度及方法	LZW; JPEG, compression level 10 (Corel Photopaint)
來源 (SOURCE)	描述數位化資料來源的實體特徵，如尺寸、狀況、原件或複本，如果為達到更佳的數位化效果，而對來源做過修正，這些修正資訊也要予以記錄。如果來源是照片或微縮型態，影像型態也要記錄(例如，正片或負片)。	照片，20 x 25 公分；原件；表面有水漬；18 x 22 公分
色彩(COLOR)	指出像素深度	1-bit ; 8-bit
色彩管理 (COLOR MANAGEMENT)	標明用以改善影像檔在掃描、呈現及輸出時色差問題的系統	Photo CD ; Optical (color management system) ; Profile/80(color sync profile maker)
色階／灰階 (COLOR BAR/GRAY SCALE BAR)	指出有無利用色階或灰階方式，並標明型態	Kodak Q13 or Q14 Color Separation Guide and Gray Scale ; Kodak Q60 Color Input Target
控制標的 (CONTROL TARGETS)	指出影像檔中用以進行品質管制、認證等標的之相關資訊	AIIM 2 號掃描測試圖；RIT 文數字解析度測試物件；

RLG 保存性後設資料的發展原是為了補 Dublin Core 及 USMARC 的不足，但該保存性後設資料僅合適影像檔。2000 年 3 月，美國線上圖書館

中心 (Online Computer Library Center, 簡稱 OCLC) 與 RLG 宣佈合作鼓勵數位資料安全又長期保存標準發展。2001-2002 年, 美國非營利性組織機構 OCLC 和 RLG 組成國際專家工作小組發展支援數位資料的長期保存後設資料標準, 提出《A Metadata Framework to Support the Preservation of Digital Objects》報告說明後設資料元素之雛型。

2003 年 6 月, 集合全球的圖書館、博物館、檔案館、政府及私人單位的數位保存專家, 組成「保存性後設資料: 發展策略工作小組」(PREservation Metadata: Implementation Strategies Working Group, 簡稱 PREMIS), 探討有關保存性後設資料標準的應用問題, 發展核心後設資料元素資料辭典, 以應用於典藏的物件, 提供保存系統中的後設資料集參考, 期使該資料辭典能夠處理更多不同類型的數位檔案。經過幾年的發展與努力, 終在 2005 年 5 月提出《Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group》報告, 內容主分三部分: (1)PREMIS 資料字典 1.0 版(PREMIS Data Dictionary Version 1.0), 介紹在長期保存應用上重要的後設資料。(2)PREMIS 工作小組最終報告(PREMIS Working Group Final Report), 提出工作小組建議的物件資訊, 討論特殊的語義單位(Semantic Units)、主題(Special Topics)與詞彙表(Glossary)。(3)資料辭典的使用: PREMIS 範例(Use of the Data Dictionary: PREMIS Examples), 圖解資料辭典中討論的數位資料與保存脈絡情境(Preservation Contexts)之應用狀況。(註52)

PREMIS 資料辭典依據 OAIS 參考模式 (Open Archival Information System reference model, 簡稱 OAIS) ISO 14721 標準建立, OAIS 資訊模型提供資訊物件和套件的分類法概念基礎架構, 可用於典藏物件和其相關連的後設資料結構。PREMIS 係由五個涉及數位保存活動的實體類型, **智慧實體、權限、物件、保存仲介與事件**發展其後設資料元素, 五個實體關連圖如圖 5-7 所示:

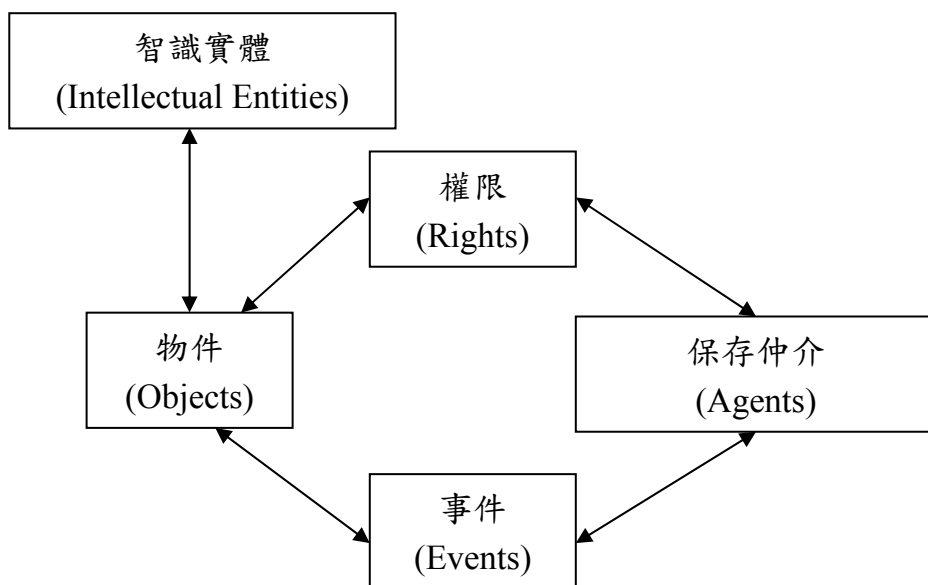


圖 5-7 PREMIS 架構實體關連圖

圖 5-7 以方框代表各個實體，以線段代表實體與實體間的關連性。箭頭的方線則代表資料辭典中定義的關連性連結方向。以下分別說明五個實體的意義：

1. 智識實體(Intellectual Entities)：是內容的集合體，可視為一個單元。例如：一本書、一張地圖、照片或一個資料庫；一個智識實體可以包含其他智識實體。例如：一個網站包含一個網頁，一個網頁又包含一張照片等。總而言之，一個智識實體可能包含一個或多個不同的數位表徵。

智識實體與物件之間的關連性於下面的例子中說明：

有一本出版於 1902 年的書，書名為《Animal Antics》，某圖書館將該本書數位化，此書共有 189 頁，為每頁各建立一個 TIFF 檔案。如結構化後設資料一般建立 XML 檔，以呈現影像與書的位置與關係。然後，圖書館會在 TIFF 影像上形成光學字元辨識(Optical Character Recognition，簡稱 OCR)。最後，用 SGML 建立文字檔，並將 189 個 TIFF 檔、一個 XML 檔以及一個 SGML 檔至於保存典藏庫中。對於典藏庫而言，《Animal Antics》這本書是一個智識實體，加上作者、題名

與出版日期等屬性就成為一個完整的單元。典藏庫有兩個表徵，一個包含 189 個 TIFF 檔和一個 XML 檔，而另一個則由一個 SGML 檔所組成。每一個表徵(Representations)使《Animal Antics》一書成為一個完整的版本。典藏庫將會紀錄有關於兩個表徵物件的後設資料與 191 個檔案物件。參見圖 5-8 智識實體與物件表徵關連圖。

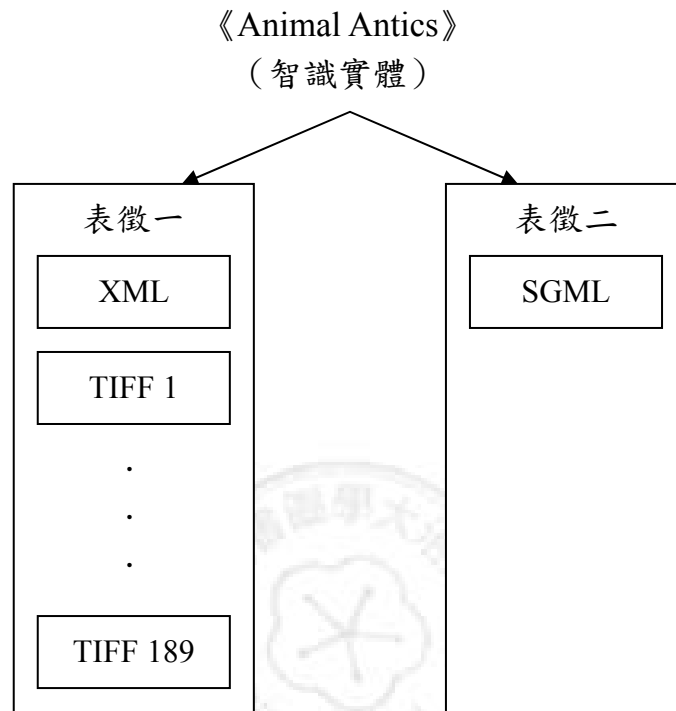


圖 5-8 智識實體與物件表徵關連圖

2. 物件(Objects)：或指數位物件(Digital Objects)，是資訊存在於數位形式之中的具體單元。物件實體有 3 個次類型：檔案(file)、位元流(Bitstream)與表徵(Representation)。「檔案」是給物件命名與順序，一個檔案可能有 0 或多個位元組，並且有一個檔案格式、取用權限(Access Permissions)、以及檔案系統統計(大小與最近修改日期)；「位元流」是檔案內連續或非連續的資料，具有保存用途的一般屬性，若沒有檔案結構的附件(如：headers 等)或把位元流再格式化為某些特定的檔案格式，則位元流不能轉換成為一個獨立的電腦檔(Standalone File)；「表徵」是指檔案的集合，包括結構性後設資料(完整與合理解釋智識實體所需)，例如：期刊論文可能完整地以 PDF 檔案格式呈現，此單一檔案即構成了物件的表徵；又其他期刊論文可能由一個 SGML 檔和兩個影像檔所組成，這三個檔案則構成物件的表徵。

3. 事件(Events)：一個事件即為一個活動(Action)，與保存典藏庫至少一個物件或機構有關。
4. 保存仲介(Agents)：保存仲介可以是一個人、一個組織或軟體程式。係指與物件生命週期內的保存事件相關的所有人、組織或軟體程式等。在資料模型圖中顯示保存仲介實體與事件實體的關連性，但保存仲介實體與物件實體則無直接關連性，必須透過事件實體間接與物件產生關連。每一個事件有一或多個相關的物件與一或多個相關的保存仲介，單一保存仲介視不同的事件可形成不同的角色。
5. 權限(Rights)：或指權限說明(Rights Statements)，是有關一個物件或仲介之一個或多個權限(Permissions)的維護(Assertions)。權限係指權利擁有者與典藏機關之間的認可，允許典藏機構採取某些行動；另外，如果權利擁有者允許典藏機構製作 3 個備份複本，這樣的描述可以視為一種權限說明，所以該權限=製作 3 個備份複本。

若以一個完整的網站快拍檔為例，其物件保存性後設資料欄位可能會有：物件識別、物件分類(Object Category)、物件特徵、Creating Application、原始名稱(Source Name)、儲存(Storage)、環境(Environment)（指軟體、操作介面、應用程式等相關之說明）、簽章資訊、關連、連結事件實體識別(Linking Event Entity Identifier)、連結智慧實體識別(Linking Intellectual Entity Identifier)等欄。參見附錄五、MINERVA 網站保存性後設資料範例。

### 三、數位典藏庫

美國國會圖書館的數位典藏庫現況，係由資訊科技服務組(ITS)負責，管理電腦儲存、電信通訊設備、系統或設備之安全、檔案與系統備份、以及所有軟硬體的操作等。為了確保在磁碟或磁帶上的數位位元之完整性，ITS 也負責管理數位內容，不論是美國國會圖書館本身所產生的數位內容，也包括從外部來源所接收的數位內容。ITS 也需要維護一個健全的儲存範圍網路，提供一個時性的電腦設備、備份、媒體轉置計畫、標準程序以及安全的政策。

美國國會圖書館與 NDIIPP 計畫係根據 OAIS 找出其需求並建立模型，現管理超過 70TB 以上的數位內容，包含：影像、網頁、照片、聲音、

影音和其他不同的數位格式及類型之數位物件。提供這些數位內容的資源提供者必須與 ITS 協調，提供有關於數位內容的 Metadata 與其他有助於檢索與使用之資訊。(註53)

從數位典藏庫系統發展的觀點來看，美國國會圖書館與 NDIIPP 計畫伙伴正進行數個典藏庫測試階段，包括：DSpace(Digital Space)、FEDORA、aDORe(a modular, standards-based Digital Object Repository)、LOCKSS(Lots of Copies keep Stuff Safe)以及 OCLC 數位典藏。(註54) 未來，國會圖書館期望多功能的典藏庫可以支援數位內容的管理、保存、與取用，並發展進階的數位保存系統，藉由科技之發展使其符合未來更多不同類型數位內容之需求。

## 第四節 永久取用

### 一、提供服務

國會圖書館提供所有數位內容的儲存、描述和取用，企圖建立不同的數位物件及與其相關的 Metadata 之長期保存。為了能夠使數位物件提供長久保存與永久取用之目的，在資源保存、互通、描述等方面，美國國會圖書館支援了多種不同 Metadata 架構的發展，採用異質性互通後設資料、描述性後設資料、權威性後設資料、以及保存性後設資料等。在未來，更將提供下列幾種系統功能：

1. 使系統多工複雜的 ingest 程序要與不同內容提供者的協議一致化。
2. 提供轉置、模擬等保存管理之功能，以克服技術/系統老化之問題。
3. 產生詳細的保存與取用報告。
4. 提供無縫式(Seamless)的取用分散式館藏與其複本。

此外，還要提供安全的儲存環境、適當取用有限制條件或有著作權保護之內容、以及走向自動化的工作流程管理。



## 二、提供取用

目前，網路資源透過 MINERVA 計畫而保存，並提供一個取用的網站，網址如下：<http://lcweb2.loc.gov/cocoon/minerva/html/minerva-home.html>。由於 MINERVA 計畫係由 17 個主題子計畫所組成，故取用介面係以主題子計畫為分類，如欲查詢與瀏覽有關 911 恐怖份子攻擊事件相關的網站資源，則需點選首頁上該主題之連結，才能進入該主題子計畫檢索或瀏覽的頁面。因此，理論上而言，網站應提供 17 個主題子計畫的連結，然而，目前可以透過 MINERVA 首頁而取用的主題子計畫僅有四個，分別是：第 107 屆國會、2002 年美國選舉、911 恐怖份子攻擊事件、以及 2000 年美國選舉雛形系統，其他則尚未建置或未開放取用。此四個主題子計畫的檢索介面係由 IA 與 WebArchivist.org 所發展，並由國會圖書館策略計畫辦公室 (OSI) 之資訊科技服務部 (ITS) 負責線上與離線取用的控制機制。

子計畫--第 107 屆國會蒐集範圍與蒐集內容為國會會員與委員會的官方網站，其提供欄位檢索與瀏覽方式以取用網站資源。在欄位檢索的部分，以輸入關鍵字檢索為主，有四種方式簡易查詢：(1)以地名為關鍵字查詢：若輸入【Texas】，則系統會搜尋與【Texas】相關的會員/委員會網站。(2)以參議員名為關鍵字查詢：若輸入【inouye】，則系統會搜尋參議員 Inouye 網站。(3)在不確定正確地名名稱的情況下，輸入其他關鍵字如“committee”，系統將會搜尋所有 committee 的網站。(4)若是在不確定參議員正確姓名的情況下，如輸入“senator Dianne”，系統也能夠精確地根據此查詢片語找到對應的結果，同時依照相關性由高至低排列，並顯示檢索結果筆數與其書目紀錄。在瀏覽功能的部分，可以國會代表或參議員人名、委員會名稱（又分白宮委員會與參議院委員會）、或者以州名瀏覽。均以英文字母 A-Z 排列。

不論以欄位檢索或瀏覽的方式取用網站資源，網站書目紀錄呈現方式均如下表 5-11 所示：

表 5-11 MINERVA 網站典藏書目記錄範例

<b>Title:</b> Julia Carson, Representing Indiana's Tenth Congressional District
<b>Alternative Title:</b> Rep. Julia Carson Website
<b>Name(s):</b> Carson, Julia

**Abstract:** Official Web site of Julia Carson, Democratic House representative from the state of Indiana, Tenth Congressional District, archived by the Library of Congress on December 12, 2002

**OriginInfo:** Washington, D.C. : United States House of Representatives, 2002

**Date Captured:** December 12, 2002

[Archived Site](#)

**Subject(s):** Carson, Julia, 1938-  
Legislators--United States  
United States--Politics and government--2001-  
United States. Congress. House  
Indiana--10th District--Democrat

**LC Classification:** E840.8.C375

**Language(s):** English

**Genre:** Web site

**Access Condition:** None

**Active Site:** <http://www.juliacarson.house.gov/>

**RecordID:** mrva0002.0149

**Collection Title:** [107th Congress Web Archive](#)

由於 MINERVA 計畫係以 MODS 作為其網站資源的描述性後設資料，故檢索結果所呈現的書目紀錄欄位大致上包括：題名、暫時題名、參議員姓名、摘要、原始資訊、擷取日期（提供當時所典藏的網站連結）、主題、美國國會圖書館館藏分類、語文、類型、取用限制、網址 URLs（動態網址）、紀錄 ID、館藏題名等。然而各主題子計畫所使用到的 MODS 欄位仍會有些許差異。

子計畫--2002 年美國選舉網站典藏，係從 2002 年 7 月開始蒐集，至 2002 年結束擷取，蒐集內容為與國會、州長、總統大選；政黨；政府；提倡；部落格；大眾見以及各種與 2002 年美國選舉相關的網站資源。此主題子計畫僅提供瀏覽功能，而無提供欄位檢索查詢的功能。瀏覽選單中有

八種瀏覽的方式分別可從選舉候選人、美國公民、Civil & Advocacy、政府、政黨、新聞報社(Press)、公共意見以及其他各式各樣(Miscellaneous)等分類瀏覽。

子計畫--911 恐怖攻擊事件，係從 2001 年 9 月開始蒐集，至 2001 年 12 月結束擷取，蒐集內容為與 2001 年 9 月 11 日恐怖組織攻擊事件發生後全球相關的報導，蒐集範圍則包括個人與團體的網站、美國新聞報刊(press)及機構等。其取用功能與 2002 年美國選舉網站典藏相同，均只有提供瀏覽功能，而無欄位查檢功能。瀏覽選單中以網站產生者、產生者類型、產生者國家、語言與主題等 5 種分類方式瀏覽。

子計畫--2000 年選舉網站典藏計畫，從 2000 年 8 月開始蒐集至 2001 年 1 月停止擷取，蒐集內容為與美國國會、州長、總統選舉；政黨；政府；提倡；大眾意見；以及各種與 2000 年美國選舉相關的網站資源。係 IA 所做的一個雛形系統，其取用介面不同於上述三個主題子計畫，其查詢功能將連結至 IA 的「回到過去」(Wayback Machine)網站取用工具，網址為：<http://web.archive.org/collections/e2k.html>，Wayback Machine 是一個可以檢索過去網站樣貌的取用工具，能夠完整地檢索從 1996 年以來的網站內容，只要輸入網址的 URL 連結，即可還原網站當時的樣貌與呈現當時既有的內容。透過這樣一個取用工具，2000 年選舉網站典藏計畫，蒐集了 2000 年美國總統大選候選人的官方網站 (George W. Bush、Gore Lieberman、Ralph Nader、自由黨候選人 Harry Browne 等)；選舉相關新聞網站 (CNN、華盛頓時報、雅虎奇摩新聞)；以及政黨網站 (共和黨)，並建置網站形成一個 URL，往後只要利用 Wayback Machine 工具，輸入 2000 年選舉網站典藏計畫的 URL，則所有當時紀錄的網站資料便得以保存與取用。

四個主題子計畫並非完全都有提供查詢檢索欄位功能與瀏覽功能。提供欄位查詢功能之計畫為「第 107 屆國會網站典藏」與「2000 年選舉網站典藏」；提供瀏覽功能之計畫為「第 107 屆國會網站典藏」、「2002 年選舉網站典藏」以及「911 恐怖攻擊事件」3 個，僅有「第 107 屆國會網站典藏」同時提供查詢與瀏覽的功能。其提供取用網站資源的方式，參見表 5-12。

表 5-12 MINERVA 四個計畫提供查詢與瀏覽功能之情形

主題子計畫	查詢功能	瀏覽功能
1. 第 107 屆國會網站典藏	●	●

2.	2002 年選舉網站典藏		●
3.	911 恐怖攻擊事件		●
4.	2000 年選舉網站典藏	●	

※ 註：有“●”符號者代表有提供該功能。

### 三、數位保存策略與技術

美國國會圖書館的數位資源長期保存策略有三個目標：(1)位元的保存；(2)內容的保存；與(3)經驗的保存(Preservation of Experience)。位元的保存係指保存原件的位元組(Sequences of Bits)；內容的保存則是保存文字內容或影像外觀；經驗的保存則指保存與數位資料的互動，包括外觀與感受(Look and Feel)、動態元素的執行等。

儘管過去幾年中，對於保存有相當深入的研究，但是在保存數位資源方面，為了完整呈現其特徵，因此大多採用的保存技術與方法是「更新」與「轉置」(Migration)。而對於網站資源的保存則採「更新快拍檔」(Refreshing the Snapshot Files)與「轉置」兩種保存技術方法。在更新快拍檔的部分，美國國會圖書館需要保留數位內容的原件與未修正過的快拍檔，就是精確地把其位元組下載起來，而這些位元組將是未來研究的主要來源資料。此外，保存原件、快拍檔亦同時達成數位保存的首要目的--「位元的保存」。在轉置策略的部分，由於電腦格式與系統不斷地在改變，10年前的檔案格式保存到現在可能已經難以完整讀取。一旦電腦系統改變，除了更新策略之外，則必須採取轉置策略，如此一來，數位資源才能夠繼續被解釋與說明，也才能透過系統與軟體的辨識、讀取瞭解資源之類型、結構與格式，美國國會圖書館的 MARC 編目記錄就是一個典型的例子。轉置的原則就在於不論資料的格式或結構如何改變，其基本的內容及語義不變。

而對於網站資料來說，自動化的轉置能夠依據檔案類型而處理各別的檔案，是可以獨立進行轉置工作的。保存系統將會檢視資料的格式、協定、程式原件等，如果檔案類型趨於快被淘汰的命運，則保存系統會自動將檔案轉換(Convert)至最新且最接近的格式。例如：一個 HTML 3.2 版的檔案，系統會轉換成 HTML 4.0 版，或者更久以後會轉置為 XHTML；又如，Java 1.0 程式可能被轉換成 Java 2.0 程式。理論上而言，自動化轉置將保存數位內容，有時還可能保存經驗；實務上而言，沒有任何轉換過程會是完美的，尤其對於可執行碼(Executable Code)或當資料可能包含了錯誤資訊的情況

之下。因此，保留物件的原始來源與轉換的紀錄是相當重要的，不僅要保存經過轉置過後的版本，最初的版本同樣也必須保存起來。(註55)

#### 四、軟體與系統

美國國會圖書館 MINERVA 網站典藏計畫，在網站徵集的部分，採用 Internet Archives 發展之開放來源蒐集器(Open Source Harvester)「Heritrix」軟體擷取網站資源；網站資源的選擇與權限認可系統採用 2006 年 9 月紐西蘭國家圖書館公布之網站擷取工具 (Web Capture Tool，簡稱 WCT)，是一種開放來源網站蒐集管理系統(Web harvesting management system)；在網站資源取用的部分，LC 採用「WERA」及「回到過去」(Wayback)兩種取用工具，以及搜尋引擎「NutchWAX」。而所有網站資源之徵集、取用與儲存的技術環境是以開放平台 Linux 架構為基礎。(註56)

#### 五、取用限制

美國國會圖書館 MINERVA 計畫蒐集的網路資源之取用，主要可區分為兩種類型：其一為開放取用的網路資料；其二為有設定取用限制條件的網站資料。而美國國會圖書館關注的則是第一種類型的網站資料，亦即提供開放取用的網站資料。開放取用資料係指網站建立者將其作品提供大眾可檢索取得，而無設定任何權限限制（如取用時間、取用地點等），圖書館能夠很容易地從網際網路上下載網頁與蒐集，然而有些資料則是可以從網路上取得卻仍有些許取用的限制，最常見的限制乃因為有機器人排除(Robot Exclusion)的緣故，亦即資料可閱讀，但可能因為網站著作權擁有者未同意下載，使讀者無法自行下載之。(註57)

### 第五節 從歷史與成功策略歸納促成因素

研究者從美國國會圖書館數位保存計畫中，歸納出下列幾點促成美國國會圖書館數位保存得以發展之因素：

#### 一、著作權法促成國會圖書館擔負起法定寄存的責任

1976 年著作權法提出在符合某些條件的情況下，為了保存或替換的目的，得以允許圖書館重製創作者的著作，其第 106 節〈Exclusive rights in

copyrighted works) 規定著作權擁有者必須：(1)複製作品複本或唱盤(phonorecords)；(2)根據該作品籌備衍生作品(To prepare derivative works based upon the work)；(3)貢獻作品的複本或唱盤予大眾，透過販賣、所有權的轉換、或出租等方式；(4)履行作品公開的原則，作品包括：文學作品、音樂劇、戲劇作品、芭蕾舞編舞(choreographic)作品、手勢語言、動作片與其他音訊作品；(5)公開展示作品；(6)聲音紀錄則透過數位聲音傳送的方式履行作品公開的原則。(註58)美國著作權法的執行機構為著作權局，其掌管各種圖書資料之著作權登記，管理美國著作權法以及有關於寄存的活動。而著作權局又隸屬於美國國會圖書館，故促成國會圖書館單負起法定寄存責任的主要國家級機關單位。

## 二、 國會圖書館的角色促成其擔負起數位保存計畫的責任

美國國會圖書館原只是單純提供立法參考服務的議會圖書館，後來逐漸發展成為全國學術界、圖書館界與一般民眾的國家圖書館，兼具議會圖書館與國家圖書館雙重角色。其任務在使其館藏資源能夠服務國會及其公民，擔負支援與保存全球文化與知識之責，提供下一代創新思考力與學習能力。然而，數位保存牽涉國家政治、法律、經濟、教育、技術等層面許多複雜的問題，絕非任何一個機構能夠獨力完成，其複雜的程度需由國家層級的單位負責，才得以妥善達到與各界溝通合作、發展的目的。美國國會圖書館向來以服務美國國會、學者專家、研究者、美國公民等之資訊需求為宗旨，也是過去法定寄存法規定出版品必須呈繳複本予以典藏及保存的資訊服務機構機構，也因此，國會圖書館替人類保存了許多文化智識的遺產，應當擔負起數位保存之責任。

## 三、 美國國會圖書館重視數位資源法定寄存問題

美國國會圖書館透過著作權法，而有責任保存作品最好的版本至少一種以上的形式，擔負法定寄存之責。但是面臨數位時代來臨，數位保存日益重要，美國對於線上出版品或是資料庫並未訂定保存版本的管理規則而無法依法使線上出版品的作者自動呈繳其作品予以圖書館典藏，是以國會圖書館在徵集數位資源以為世代永久典藏與取用時，面臨到一個典藏機關與創作者間寄存共識的無形屏障。因此，國會圖書館著作權局乃著手進行有關於電子著作權註冊和寄存的先導計畫。隨後於 1988 年，國會圖書館開始要求寄存磁帶、光碟和微縮資料。1993 年，建立了 CD-ROMs 的自願

寄存協議。2001年，第一本電子書透過電子著作權法而開始寄存，但至今仍無有關於強制或義務寄存線上電子出版品的寄存法與規定。因此，目前在蒐集與擷取數位資源產生複本的同時，必須先徵得產生者的同意。

#### 四、 美國記憶與國家數位圖書館計畫成為先導

美國國會圖書館自從1994年執行國家數位圖書館計畫(National Digital Library Program, 簡稱NDLP)以來，最主要的其中一項成果即為美國記憶計畫(American Memory Project)，該計畫是一個大型數位化計畫，使美國國會圖書館的館藏可以透過網際網路提供全球的使用者取用，獲得全國甚至全球使用者的支持與喜愛，廣受好評。而NDLP計畫中所建立的數位化過程、永久辨識號(Persistent Identification)與儲存管理(Storage Management)之實務也廣泛地被應用到其他有關數位內容取用的計畫，例如：全球入口閘道(Global Gateway)、音樂領域的數位館藏、印刷品與照片的線上編目、退役軍人歷史計畫(Veterans History Project)。(註59)有了美國記憶計畫與國家數位圖書館計畫之成功經驗與成果，促成美國國會圖書館繼續朝向發展國家數位資訊基礎建設與保存計畫，MINERVA計畫等數位保存活動。

#### 五、 國會圖書館行政體系促成其較易於爭取各項發展經費

國會圖書館組織龐大、體系複雜，其行政體系與圖書館界的地位隸屬於美國國會，故較為能夠受惠於國會通過各種相關立法案，進行擴展規模計畫，爭取到相關經費。

#### 六、 重視數位保存觀念

促成美國開始探討數位保存議題之因，是由於數位時代來臨，數位資源大量以數位形式增加，且越來越多數位資訊內容重要卻沒有任何實體形式的資料格式存在。今日所檢索取用的資料，可能在明日便無法再次取用，根據調查，網站壽命平均只有44天，網站連結失效、內容遺失、媒體格式改變、軟體程式更新、硬體設備更新而無法支援舊版，使得重要的數位資源無法辨識讀取，均造成社會大眾你我的損失，於是各種資訊服務機構、資訊提供者、出版者、資訊內容擁有者、學者專家、社會大眾等便開始注意到數位保存的重要性，逐漸關注於數位保存的問題。

## 七、 面對數位保存挑戰產生數位保存計畫構想

促成美國國家圖書館數位保存計畫發展除了大家開始關心數位資源保存的議題之外，圖書館為了要能提供豐富的館藏資源與無時差、無距離的館藏取用，以滿足使用者資訊需求之目的，乃促成數位化資源典藏計畫與數位圖書館計畫的開始，因為豐富的館藏需要數位化才能達到加強館藏取用之目的，而數位典藏計畫亦間接地影響了日後原生數位資源保存計畫之發展。有了美國記憶數位典藏先導計畫、以及數位圖書館計畫等均深受國人喜愛，這些數位典藏與數位圖書館計畫，將實體館藏數位化，促使其館藏資源更加透明化，不僅將實體以數位化的方式保存，更增加其使用性與瀏覽性。更是替數位保存計畫早先奠定了基礎與建構未來數位保存技術架構、標準和政策、策略等發展。

## 八、 訂定數位保存策略目標

美國國會圖書館發展「國家數位資訊基礎建設與保存計畫」(NDIIPP)，設定五項策略性目標：(1)辨識與選擇具風險性的原生數位內容；(2)建立與支援數位保存伙伴網；(3)發展與應用在數位保存用途之工具、模型、方法和服務；(4)鼓勵並支援數位保存公開政策；(5)傳播數位保存重要的觀念。

## 九、 重視網路資源的保存與典藏

美國 OCLC 調查網站生命週期的狀況，發現網站資料穩定地不斷成長，但網站消逝的速度亦十分快速。根據統計，網頁平均壽命僅有 44 天。國會圖書館為使其資源能夠被國會會員與美國人民長久取用、可及，以支援全球知識館藏，培養未來世代子孫的創造力與學習，乃於 2000 年正式發展「網際網路電子資源虛擬典藏計畫」(MINERVA)。在龐大而複雜的網路世界中，建立有關於美國政治、經濟、社會、教育、科技等層面相關的主題網站典藏計畫，以保存人類文化知識的遺產。

## 十、 明確的數位保存館藏政策與選擇標準

美國國會圖書館為使網路資源選擇與蒐集有依循的標準與規範，產生「電子資源館藏發展說明」、「網站擷取與典藏」、「網站擷取與典藏特別指南」、以及「Questions for Selectors for BEOnline Plus」四種有關於電子資



源館藏發展與選擇標準的參考文件。強調蒐集與選擇網站的人員必須考量數位資源的取用性、權威性、內容正確性、使用者介面與導航性、獨特性，判斷是否為具有遺失風險的數位內容、資料內容是否即時且適用、是否具有研究價值與貢獻等問題。

## 十一、 採取合作蒐集數位資源的策略，建構國家伙伴網

數位資源保存相當複雜與困難，包括數位資源的辨識、蒐集、保存、維護、提供取用等工作，非任何單一機構或組織能夠獨立完成，就連全世界規模最大、館藏最豐富的美國國會圖書館亦不例外。故國會圖書館在面臨數位資源保存的問題時，著重採取合作的方式從公私領域相關組織及機構尋求合作伙伴，共同研究數位保存相關政策、標準與技術等，並蒐集與保存對國會與公民有用且相關的數位資訊。其中 NDIIPP 計畫的策略之一，即是與國內大學圖書館、博物館、檔案館等資訊服務機構，甚至擴及國際間網際網路聯盟(IA)、國際網際網路保存聯盟(IIPC)等建立合作伙伴聯絡網，朝向數位保存研究、數位保存技術工具與架構、辨別與選擇具風險性的原生數位內容等目標策略一一達成與實踐。

## 十二、 免費的擷取網站工具

國會圖書館採用 HTTrack 軟體來下載網站快照檔案，其可爬梳網站中所有檔案與連結，下載網站中的網頁直到整個網站都已被複製起來，包括文字檔、影像檔、影音檔、執行程式檔、後設資料檔與樣式檔(Style Sheets)等。

## 十三、 實際採取數位保存技術策略

美國採用更新網站快照檔與轉置策略，使數位資源得以延長生命週期進而長期保存，使其能提供永久取用的服務。

## 十四、 訂定合適的後設資料標準與架構

美國國會圖書館支援多種不同 Metadata 架構，包括：後設資料編碼與傳遞標準、後設資料物件描述標準、後設資料權威描述標準、以及 Metadata for Images in XMLMIX。並採用「PREMIS 資料辭典」作為其保存性後設

資料標準。(註60)

值得一提的是，美國國會圖書館採用 MODS 作為其網路資源的描述性後設資料。原因是 MODS 可用於各種書目元素集的架構，也可將圖書館與現有的 MARC 格式相互轉換，並以 XML Schema 作為文件結構，兼容現有的 MARC21 紀錄資料，並可建置原始資源的書目描述紀錄。

此外，PREMIS 的發展也帶給美國國會圖書館在數位資源保存性後設資料方面之參考標準。PREMIS 可以支援更多類型數位資料的保存性後設資料描述，而非如 RLG 於 1996 年所設計的 16 項保存性後設資料欄位，僅合適於描述影像檔案的保存性後設資料。

## 註釋

---

- 註1 Library of Congress, “Home Page,” <<http://www.loc.gov/>> (retrieved April 4, 2007).
- 註2 李淑玲，美英兩國國家圖書館體制與功能之比較研究（臺北市：漢美，民國 79 年），頁 25-28。
- 註3 Library of Congress, “Financial, Annual Reports,” <<http://www.loc.gov/about/reports/>> (Retrieved at May 22, 2007).
- 註4 United States Copyright Office, 2004-2005: Strategic Plan 2004-2008. United States Copyright Office.
- 註5 國立編譯館，圖書館學與資訊科學大辭典(中冊)（臺北市：漢美，民國 84 年 12 月），頁 1516。
- 註6 Copyright Office, “What Is Copyright,” <<http://www.copyright.gov/circs/circ1.html#wci>> (retrieved May 4, 2007)
- 註7 Ingeborg Verheul, Networking for Digital Preservation: Current Practice in 15 National Libraries(Munich : K.G. Saur, 2006), p.207.
- 註8 Copyright Office, “What Is Copyright,” <<http://www.copyright.gov/circs/circ1.html#wci>> (retrieved May 4, 2007)
- 註9 同註 7。
- 註10 Library of Congress, “Importance of Digital Preservation,”

- 
- <<http://www.digitalpreservation.gov/importance/>> (Retrieved at May 10, 2007).
- 註11 Library of Congress, “About the Digital Preservation Program,”  
<<http://www.digitalpreservation.gov/library/about.html>> (Retrieved at June 9, 2007).
- 註12 Library of Congress, “What the Library Is Doing – Meeting the Challenge: Saving the World Wide Web,”  
<<http://www.digitalpreservation.gov/library/>> (Retrieved at May 16, 2007).
- 註13 同註 12。
- 註14 Library of Congress, “What the Library Is Doing - What Is Being Saved,” <<http://www.digitalpreservation.gov/library/collect.html>> (Retrieved at April 23, 2007).
- 註15 同註 14。
- 註16 同註 14。
- 註17 同註 14。
- 註18 同註 14。
- 註19 Library of Congress, “Partners,”  
<<http://www.digitalpreservation.gov/partners/>> (Retrieved at May 17, 2007).
- 註20 Library of Congress, “What the Library Is Doing - Home,”  
<<http://www.digitalpreservation.gov/library/>> (Retrieved at May 16, 2007).
- 註21 同註 7。
- 註22 Library of Congress, “Achievements,”  
<<http://www.digitalpreservation.gov/achievements/>> (Retrieved at May 10, 2007).
- 註23 同註 7，頁 219-222。
- 註24 同註 7，頁 222-223。
- 註25 同註 7，頁 223。
- 註26 Library of Congress, “Web Capture,”  
<<http://www.loc.gov/webcapture/faq.html>> (Retrieved at April 16, 2007).
- 註27 Library of Congress, “Web Capture & Archiving,”

- 
- <<http://www.loc.gov/acq/devpol/webarchive.html>> (Retrieved at April 16, 2007).
- 註28 “About MINERVA,”  
<<http://lcweb2.loc.gov/cocoon/minerva/html/minerva-about.html>>  
(Retrieved at April 16, 2007).
- 註29 Library of Congress, “Web Capture – Partners,”  
<<http://www.loc.gov/webcapture/partners.html>> (Retrieved at April 16, 2007).
- 註30 Diane Kresh, “Harnessing the Web: The MINERVA Program at the Library of Congress,” 25 March, 2003,  
<[http://www.loc.gov/minerva/presentations/diane\\_rlg.ppt](http://www.loc.gov/minerva/presentations/diane_rlg.ppt)> (Retrieved at April 16, 2007)
- 註31 Library of Congress, “Web Capture - Projects,”  
<<http://www.loc.gov/webcapture/projects.html>> (Retrieved at April 16, 2007).
- 註32 Library of Congress, “Web Capture,”  
<<http://www.loc.gov/webcapture/index.html>> (Retrieved at April 23, 2007).
- 註33 同註 12。
- 註34 同註 7，頁 208。
- 註35 同註 7，頁 208。
- 註36 同註 7，頁 208。
- 註37 Library of Congress. “What the Library Is Doing - Library of Congress Digital Executive Oversight Group.”  
<<http://www.digitalpreservation.gov/library/oversight.html>> (Retrieved at May 16, 2007).
- 註38 同註 7，頁 210。
- 註39 同註 14。
- 註40 Library of Congress, “Electronic Resources (1999),”  
<<http://www.loc.gov/acq/devpol/electron.html>> (Retrieved at May 23, 2007).
- 註41 Library of Congress, “Electronic Resources, Selection Guidelines (2004),”  
<<http://www.loc.gov/acq/devpol/electronicselectionguidelines.html>>  
(Retrieved at May 23, 2007).

- 
- 註42 Library of Congress, “Web Capture & Archiving,”  
<<http://www.loc.gov/acq/devpol/webarchive.html>> (Retrieved at May 23, 2007).
- 註43 BEOnline+ Project, “Selection Criteria for Internet Resources included in BEOnline Plus,”  
<<http://www.loc.gov/rr/business/beonline/beonsel.html>> (Retrieved at May 23, 2007).
- 註44 “Questions for Selectors for BEOnline+,”  
<<http://www.loc.gov/rr/business/beonline/questions.html>> (Retrieved at May 23, 2007).
- 註45 MINERVA, “September 11 Web Archive,”  
<<http://lcweb2.loc.gov/cocoon/minerva/html/sept11/about-selection.html>> (Retrieved at May 17, 2007).
- 註46 William Y. Arms, Roger Adkins, Cassy Ammen, and Allene Hayes, “Collecting and Preserving the Web: The Minerva Prototype,” RLG DigiNews, <<http://www.rlg.org/preserv/diginews/diginews5-2.html#feature1>> (Retrieved at May 23, 2007)
- 註47 Internet Archive, “Heritrix - Introduction,” <<http://crawler.archive.org/>> (Retrieved at May 23, 2007).
- 註48 Library of Congress, “Web Capture - Technical Background,”  
<<http://www.loc.gov/webcapture/technical.html>> (Retrieved at April 16, 2007).
- 註49 同註 46。
- 註50 MINERVA, “Descriptive Metadata Component”  
<<http://lcweb2.loc.gov/cocoon/minerva/html/minerva-metadata.html>> (Retrieved at May 31, 2007).
- 註51 同註 7，頁 213。
- 註52 OCLC, “PREMIS (PREservation Metadata: Implementation Strategies) Working Group,” <<http://www.oclc.org/research/projects/pmwg/>> (Retrieved at May 17, 2007).
- 註53 同註 7，頁 211。
- 註54 同註 7，頁 213。
- 註55 同註 46。
- 註56 同註 48。

---

註57 同註 46。

註58 Copyright Office, “What Is Copyright,”  
<<http://www.copyright.gov/circs/circ1.html#wci>> (retrieved May 4, 2007)

註59 同註 7，頁 208。

註60 同註 7，頁 213。

