

我國現代文學史料數位化典藏與服務（上）



莊健國・國家圖書館參考組編輯

一、前言

台灣地區現代文學發展歷程坎坷，一般說來，打從日據時代開始，台灣新文學已經開始萌芽，此一時期代表作家有賴和、吳濁流、蔡秋桐、王白淵、謝春木、楊守愚、楊逵、楊華、楊雲萍、吳新榮、王詩琅、郭水潭、楊熾昌、張文環、龍瑛宗、巫永福、呂赫若、林芳年、鍾理和、吳瀛濤、王昶雄等人，其作品無論是漢詩、白話詩、民謠、散文、小說等，往往傳達著濃厚的殖民文化——從宿命、鄉愁、悲情的鄉土情感到抵抗、挑戰等嚴肅主題，象徵著台灣人精神苦悶的年代。

台灣光復後，國民政府接收台灣，台灣人民正準備回歸戰後安定的日子，孰料國語的強制推行及二二八事件的歷史悲劇，使得台灣文學遭逢斷層的不幸，例如呂赫若、張文環、楊逵、王白淵等人，分別遭到

程度不一的政治迫害，有些作家因此封筆退隱，文學作品銳減。（佩謙〈淺談台灣新文學〉，《傳統中國文學電子報》第十九期，民國八十八年九月五日，<http://cyberfair.taiwanschoolnet.org/c00/21200085/w4.htm>）此後台灣地區文學發展主流以外省籍作家為主，例如散文家琦君、王鼎鈞、夏元瑜、林海音等，約至一九七〇年代本省籍作家對國語的文字處理較為成熟，逐漸進入所謂「鄉土文學」時期，此一時期文學作家不分省籍，偏好描述周遭的生活環境與景物，或為市井，或為鄉村，或為社會，就地取材，關懷各階層生活百態。

到了一九八〇年代以後，台灣本土文學再興，新的台灣現實意識也開始萌芽，潛伏的本土論者在有利的客觀環境中，開始探出再出發的腳步。八〇年代，台灣社會的歷史主體意識普遍覺醒，台灣人又開始在台灣座標尋求台灣的出路，本土論因此受到重視而得



到普遍的認同。(游勝冠《台灣本土論的興起與發展》，東吳大學中國文學研究所碩士論文，民國八十年，頁一、六)

從上述台灣現代文學的發展歷程，不難想像此一時期之文學史料不容易徵集與典藏，一方面是殖民時代距今年代久遠，資料典藏條件較差，且出版品目錄不完整，文學史料徵集實難齊全，另一方面由於政治主權轉換，有些文學家擔心政治迫害，不輕易發表作品，因此有部分作品僅剩孤本手稿，其保存條件與環境令人憂心，最後往往僅能從其遺稿中找到若干創作，而文學家家屬如果不知如何保存這些珍貴文學史料，最後恐也流落廢紙廠或垃圾堆中，殊為可惜，因此早期台灣現代文學史料之度藏，除了倚重文學家家屬外，圖書館及文學社團的參與典藏，也是非常重要的一環。

由此可知早期台灣現代文學史料的徵集與典藏都不容易，更遑論其閱覽服務，因為讀者不知道有那些史料可以參考，也不知道典藏於何處，常需耗費時間，奔波於各大圖書館或文學相關資料室，因此政府刻正進行的「台灣地區地方文獻典藏數位化」計畫是非常有意義的工作，如能儘早完成，對於各領域研究者，助益匪淺。

本文作者依其在文學史料領域多年來之實務工作，淺談台灣地區文學史料數位化經驗，先從文學史料數位化之定義說起，再敘述文學史料之發展背景，

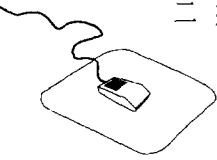
進而簡述文學史料數位化之特徵及數位化典藏服務與數位化規劃要點，最後再論及文學史料數位化發展現況與未來趨勢，尚祈各界不吝指正。

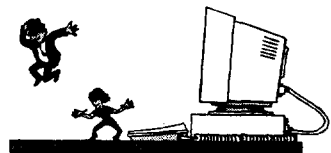
二、文學史料數位化的定義

「文學」一詞根據《正中形音義綜合大辭典》一書所載：「近世所謂文學，有廣狹二義：廣義泛指一切思想之表現，而以文字記述之者；狹義則專指偏重想像及感情之藝術作品，如詩歌、小說、戲劇等屬之，亦稱純文學。」

「史料」一詞在前書中則定義如下：「研究或編纂歷史的材料，有實務、紙面、口傳三種。」另在《中文大辭典》一書，則解釋為「史料也，歷史之參考資料也」。

「數位化」一詞在《英漢計算機百科辭典》一書之解釋如下：「將物理變數的類比量度轉換為數值，而以數位形式改變該量。簡單的說，通信系統中的數位化通常是指將語言或圖形信號變為二進碼。實際上任何形式的資料在計算機能接受它們之前都要被數位化」(林在高、宋文強主編《英漢計算機百科辭典》，台北中央圖書，民國八十五年，頁四五七、四五八)；另陳俞玟在其文章中將數位化一詞解釋為「將文字或圖像經由數位掃描器、數位照相機的轉換，而成為電腦所能處理的數位數據。」(陳俞玟《淺談資料的數位轉換》，《國立中央圖書館館刊》二





十八卷二期，民國八十四年十二月，頁三十二)

綜合上述定義，「文學史料數位化」應該可以解釋為「泛指將一切文學活動的紀錄以電腦格式儲存，包括文學創作、評論、人物傳記、圖片、手稿、聲音、視訊等各類型媒體予以數位化，然後可以再經由電腦介面將上述文件或資料以其原貌呈現。」

三、文學史料數位化的歷史背景

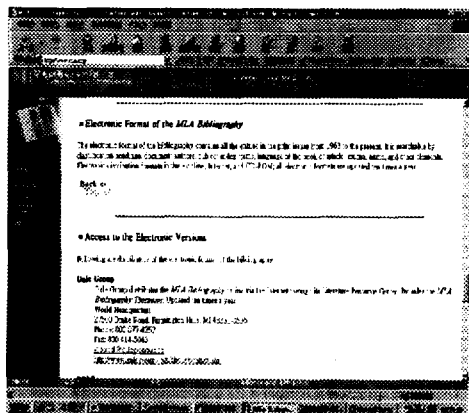
文學史料數位化的歷史比起其他領域科學似乎稍微落後一些，這是可以輕易理解的，因為國際網路發展之初，原來是以軍事用途為導向，後來才逐漸使用於科學、商業與教育等領域，但這並不減損文學史料之重要性，在國際網路蓬勃發展10年後的今日，學術界有一個共識，那就是「Contents」的充實才是全球各學科領域目標一致的方向。「資料內容(Contents)」將是未來網路持續發展的主要條件，那些內容貧瘠、資料不完整、不具權威性、專業性，或無特色的網站，將很難在下一個世代存活於學術網站領域之中。

文學史料資料庫的歷史首推「美國現代語文學會」(Modern Language Association, 簡稱MLA)的摘要系統，這個已經有40年歷史的資料庫，收錄一九六三年起，有關文學、語言及語言學、民俗學等學術領域之期刊、叢刊、專書、論文、會議論文集及書目等，編製成索引摘要資料庫(MLA International Bibliography)，以供蒐尋相關資料來源，是研究西方

語言、文學不可或缺的重要參考工具，也是語言文學領域最權威的資料庫，唯至今尚未推出全文連結服務。

「古騰堡計畫」(Project Gutenberg)是美國伊利諾大學於一九七一年開始進行，將屬於公共版權的文學資料如「愛麗斯夢遊仙境」、「莎士比亞全集」、「羅傑辭典」等全文輸入電腦，供網路上讀者閱讀。這是文學史料數位化的重要里程碑，目前全球有好幾個網站免費提供該計畫建置的數位化書籍之下載服務。

Contemporary Authors 是另一個著名的作家作品與人物評介工具書，也有超過50年的編纂歷史，是該領域重要之參考書。原本是以印刷式書本發行，目前是改以光碟版發行，收錄一九六〇至一九九六年間約10



▲美國現代語文學會 (MLA) 首頁
(<http://www.mla.org>)

萬位當代作家資料，包括小說家、詩人、劇作家、報界人士及非小說類文學作家。
Disclit 系列光碟，則包括 American Author、British Author、World Author 等文學系列叢書共約500冊之全文資料，是研

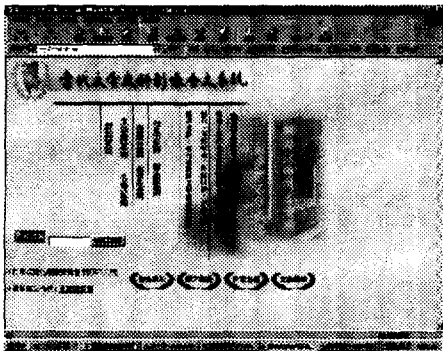




究美國文學、英國文學及世界文學之入門參考書。

Essay and General Literature Index 收錄美國、英國及加拿大地區所出版發表的英文論文及文章，重點強調在人文及社會科學。所包括的主題有：經濟、政治、歷史、戲劇、電影、文學作品的評論等。每年約收錄300冊書籍及20多種年刊及期刊中的文章。

文學類資料在我國圖書資訊界的數位化歷史，似乎略佔優勢，以國家圖書館為例，文學史料是館內第一個數位化成功的領域，其原因有四：一則原先收錄之資料已具備數位化之基礎，因為該館曾於一九八五年五月辦過一次「當代文學史料展」；二則它的學術領域範圍較明確，資料蒐集時較易掌握；再則國家圖書館擬先小規模試作數位化工程，以汲取、累積經驗，再擴大範圍至資料龐雜的期刊報紙、學位論文、善本古籍等類別資料；最後，也是最重要之因素，則是該館正好搭上「六年國建」專案計畫列車，獲得行政院的專款經費補助，使得該計畫得以順利推展成功。國家圖書館於民國一九九四年順利推出「當代文學史料影像全文系統」，確實具有先導性及指標性之重大意義，該系統收錄五十年來台灣地區當代文學作家約兩千餘位之基本資料及其生平傳記、手稿、照片、著作年表、作品



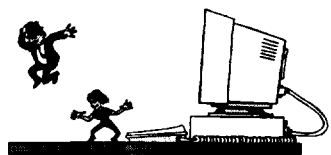
▲當代文學史料影像全文系統首頁
(<http://lit.ncl.edu.tw>)

目錄、評論文獻、翻譯文獻、名句及歷屆文學獎得獎紀錄，每年平均有20萬人次上網使用該資料庫，頗受文學界人士喜好，是國內文學網站之重鎮，對文學研究學者及熱愛文學之民眾，提供便利之檢索功。唯該系統對於文學作品之典藏尚未建置，殊為可惜，而收錄文學作家之全面性與評述作家基本資料之完整性仍待改善，如能調整上述缺憾，方能成為台灣現代文學史料的重要資源。

中央研究院自一九八四年起開始建置「漢籍全文資料庫」(<http://www.sinica.edu.tw/~tdbproj/handy1/>)，至一九九八年止先後完成「二十五史」、「十三經」、「諸子」、「古籍十八種」、「古籍三十四種」、「大正新脩大藏經」等文獻之數位化工作，超過兩千萬字的台灣史料、一千萬字的大正藏以及其他典籍，合計字數一億三千四百萬字，並以每年至少一千萬字的速率，持續成長，是當今國內最具規模的中文古籍資料庫。

元智大學的「網路展書讀」(<http://cls.admin.yzu.edu.tw/>)是以古典文學多媒體服務為目標之網站，一九九三年起即開始中國古籍數位化的研發工作，目前已完成中國古籍數位研究系統包含：《詩經》、《全唐詩》、《宋詩》、《唐宋詞》、《紅樓夢》





等，累計達一千五百萬字。已完成的網路教學系統包含：倚聲填詞格律自動檢索引教學系統、依韻入詩格律自動檢索引教學系統、唐詩三百首、網路對聯習作系統等。該網站在國際網路崛起之初即推出，又能持續建置與推廣，因此獲得不少嘉評，並對往後文學網站製作者具有催化作用。

「四庫全書電子版」是兩岸三地合作開發之電子資料庫，收錄文淵閣四庫全書共三千四百八十種典籍，三萬六千冊圖書，四百七十萬頁影像，八億漢字之中華文化巨著，四庫全書電子版以Unicode內碼儲存資料，並補充自建造字檔，方便處理與顯示中文罕用字，而其全文檢索功能更讓研究人員津津樂道，除了可節省時間之外，更不會遺漏任何一則相關訊息，也不必再像從前，往往為了某一句話或某一特定主題，翻遍了整部四庫全書，還不一定找得到資料。電子版四庫全書之售價又比原書便宜，並可節省圖書館書架空間，是學術圖書館數位化的重要資源。

「康熙字典」光碟是根據「同文書局」出版的石印版《康熙字典》為底本製作而成，共收錄四萬七千零三十五字，分為二百一十四個部首，內容極為豐富，但其編輯體例對於今日之一般讀者查詢使用卻顯得有點不便，其電子版採用先進的電腦檢索技術彌補此一不足，提供單字、部首、筆畫、筆順、注音、拼音等多元化的檢索欄位查詢，讓讀者能夠輕鬆使用此一漢學研究中重要的工具書典籍，還可透過「瀏覽原書」

的功能，直接翻閱《康熙字典》的原文圖像。

「故宮「寒泉」古典文獻全文檢索資料庫」(<http://210.69.170.100/s25/index.htm>)，由陳郁夫教授製作，將多年來自行開發之古籍文獻，免費提供學界使用，目前已完成十三經、先秦諸子、全唐詩、宋元學案、明儒學案、朱子語類、資治通鑑、廿五史等古典文籍，刻正進行古今圖書集成數位化工程。

「傳記文學」光碟收錄該雜誌第一期至第四三三期(1962.6~1998.6)間整本雜誌全文內容，共一萬五千篇文章，一億餘字數，錄製於十張光碟資料中。內容包括自傳、評傳、年譜、回憶錄、名人日記、重要手蹟、珍貴史料、史事研究等傳記文學與民國史料，提供各篇文章的日期、期號、標題、副標題、作者、專欄、人物、機關團體等項目的查詢，使用者可透過全文檢索軟體，針對任一欄位、跨欄位或所有檢索項目，輸入任意字詞進行查找，即可查獲相關書目資料並同時瀏覽內文影像。

「文訊雜誌紀念光碟」收錄該雜誌第一期至二〇〇〇期(1983.07~2002.6)之間所有雜誌的全文內容，製作於DVD光碟片之中，共約三千萬字，六千餘幅圖，提供雜誌篇目、專題、作者、被評論者、專欄、卷期的瀏覽與全文檢索功能，預計於二〇〇二年底發行。

「向陽工場」(<http://necity2.web.hinet.net/UserData/hylim/default.html>)由詩人向陽先生主持，成立於一九





九八年，收錄「向陽的詩」「台灣文學」「文學傳播」「報導文學」等文學資源，向陽先生在網際網路環境下嘗試多角經營文學網站，相關的網站還包括「向陽文苑」、「向陽詩房」、「台灣文學傳播研究室」等，是國內網路文學的重要靈魂人物。

「傳統中國文學網」(<http://www.literature.idv.tw>)乃遙光先生創設於民國一九九八年，以發行「傳統中國文學電子報」起家，經過數次改版與更換網址，並積極尋求合作對象，引進各類型文學資料，以網路行銷策略，厚植網路文學網站聲望，是近年來急速竄紅的文學網站。網站上除發行電子報外，還提供國學入門、文學研究專區、電子書庫、中文研究所考訊等資訊，是個人經營文學網站成功之典範。

「南方社區文化網路」(<http://www.esouth.org>)創立於一九九五年，發行台灣本土第一份對大眾發行的「南方電子報」，宣揚其服務宗旨——「讓商業邏輯下失去戰場的理想在網路發聲」，並以「社會運動」、「社區運動」、「生態環保」、「弱勢團體」、「文學文化」為網站五大方向。

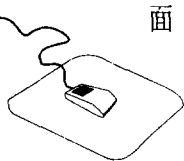
「詩路：台灣現代詩網路聯盟」(<http://www.poem.com.tw/>)由網路詩學專家須文蔚先生創立於一九九六年，是文建會長期支持的網站，集錄大量現代詩資料，是現代詩的專門網站，包括每日一詩、經典作品典藏、詩集出版簡目、重要詩集精選，以及詩學教室、理論、評論等。

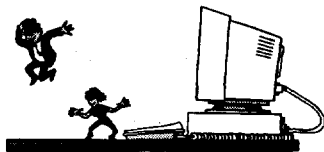
「台灣文學研究工作室」(<http://ws.twlncku.edu.tw/>)由呂興昌教授主持，對台灣地區各類型本土文學廣泛探討與推廣，網站內容除台灣文學研究學者的簡介及相關學位論文外，並有原住民文學、民間文學、母語文學、台灣古典文學及新出土的文學資料，以及鄭經時期、清領時期、日治時期及國統時期的文學研究等，並收錄相關的研究書目、研究刊物、研究計畫、學術會議及文學課程等。

「文學咖啡屋」(<http://www.cca.gov.tw/coffee>)由行政院文建會與聯合報副刊聯合策畫，提供網路創作與聯合副刊等文學作品，該網站舉辦多結局小說網路大競寫活動，每月均有不同的競寫題目，入選作品並於每月底與聯副同步刊登。

「歧路花園」(<http://audinchnu.edu.tw/~garden/garden.htm>)由中興大學外文系辦公室李順興老師主持，旨在實驗網路功能，以之作為文學創作和發行工具，嘗試開發新型的文學表現形式並改良文學傳播的管道。

台灣地區一般大學BBS站多有文藝相關版面，近年來興起的藝文業BBS站有中山大學音樂系的「山抹微雲」、清華大學的「楓橋驛站」、海洋大學「田寮別業」及成功大學的「貓咪樂園」。「尤里西斯文社」則是最早成立的藝文專業BBS站，收有許多早期BBS寫手的優秀作品。其中海洋大學的「田寮別業」則是專業藝文創作站，除了文學討論、創作的版面





外，更有「田寮書庫」收錄詩經、文心雕龍、老子到十三經、諸子百家、歷史傳記到歷代詩詞，是一個相當完整的資料庫。另外寫手們可於此版設立一個個人專屬版面，收藏自己的作品，也可與好友們討論創作問題。

大陸地區由政府部門建置的現代文學網站尚未多見，但是民間網站卻有不少全文型文學網站，例如「木子書屋」(<http://library.muzi.com/>)、「天一書樓」(<http://www.cnool.net/tianyige/>)、「網絡圖書」(<http://www.gotobook.net/>)、「亦凡公益圖書館」(<http://www.shuku.net/>)等，都可在這些網站上輕易取得部分文學作品的全文內容，唯這些網站是否已取得文字著作權上網之授權，似乎還有待確認。例如有些



▲超星數字圖書館首頁
(<http://www.ssreader.com>)

網站將台灣地區作家之作品，如瓊瑤的《庭院深深》、龍應台的《野火集》、三毛的《撒哈拉的故事》、李敖的《北京法源寺》等作

品，毫無保留的公開於網路上，雖然方便網友與讀者利用，但仍應尊重智慧財產權。這些網站大都小規模經營，既無網站設立或服務宗旨，又無網站負責機構，且非正統資料庫管理架構，較難期望網站經營的長遠性。唯有半官方「超星數字圖書館」

(<http://www.ssreader.com>) 才是有計畫有規模的經營各學科史料數位化工程，這網站是由中國國家圖書館與北京世紀超星有限公司合辦，目前已有近十餘萬冊圖書上網，全部可以在線上瀏覽(需加入會員)，號稱是大陸地區目前最大的網上圖書館。收錄的學科領域涵蓋哲學、宗教；法律、政治、軍事、經濟、教育、體育等社會學科；數學、物理、化學、天文、生物等自然科學；建築、資訊、電機等應用科學；世界史地；文學藝術等。目前的服務方式是採會員制，會員需先購買一定金額之「讀書卡」，待閱讀書刊時，系統會自動扣除儲值點數，非會員只能預覽數位化書刊之部分影像資料，約十來頁。

香港地區的文學資料庫，則以香港中文大學之「香港文學資料庫」(<http://hk.libpub.lib.cuhk.edu.hk/>)為代表，該資料庫收錄之資料以香港地區發表之現代文學創作或評論著述為主，資料收錄時間可追溯自一九三〇年代，部分資料提供全文閱讀服務，但僅限於校園內使用，是研究香港文學作家的重要資源。

以上僅略述台灣地區現代文學相關系統與網站，唯礙於篇幅，難免有遺珠之憾，尚祈見諒。