



我國現代文學史料數位化典藏與服務(下)

莊健國·國家圖書館參考組編輯

六、文學史料數位化的服務

文學史料數位化服務是未來全球資訊服務的一環，它可以結合電子商務，也可以是圖書館館內服務的延伸，也可能轉型為個人化圖書館型態，也可能結合 PDA 或手機業者的加值服務。

我們先從傳統的圖書館服務說起，以前在圖書館尚未收藏數位化資料前，讀者必須先到圖書館，利用各種目錄索引來尋找資料的出處，來後再到書架上找尋書刊，然後才能將資料複印，以便研讀。例如我們如果要找尋有關作家「林海音」的相關報導與作品評論，在還書目索引還沒有自動化之前，必須很辛苦地找尋各月份的文學雜誌與各報紙副刊及圖書目錄，才能尋獲部分資料，不但沒有效率，資料蒐集也不完備。現在我們只要在電腦前面鍵入「林海音」三個字，就可以快速的尋獲資料，甚至可以馬上閱讀雜誌

或報紙上的內容，只要該份文件已數位化掃描或建檔。

目前從事文學研究者在蒐集資料時，應該可參酌下列步驟取得目錄或全文資料：

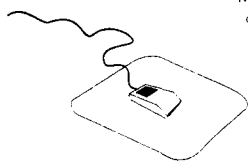
(一) 文學圖書：

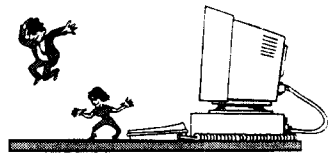
可使用「當代文學史料影像全文系統」、「國家圖書館館藏目錄系統」、「全國圖書目錄系統」，大部分的國內文學圖書，應該都會收錄在這三個系統內。

(二) 期刊文獻：

可先考慮使用「當代文學史料影像全文系統」、「中華民國期刊論文索引」、「國家圖書館期刊目次系統」、「中國期刊網」，若還不滿意，再檢索單一期刊資料庫，如「文訊光碟」、「傳記文學光碟」、「光華雜誌光碟」等，大部分的國內文學圖書，應該都會收錄在這三個系統內，且大部分都提供全文列印功能。

(三) 報紙文獻：





可先使用「當代文學史料影像全文系統」、「報紙標題索引」查詢欲檢索之人名或主題，查到後如有影像，可立即列印出來，不用再調閱舊報紙或報紙微捲資料。

(四) 學位論文：

有關文學相關論文，使用「全國博碩士論文系統」，可免費取得近幾年來已授權學生之論文全文。早期的博士論文也已數位化，但僅限在國家圖書館內使用。

(五) 研討會論文集：

有關文學研討會的論文集較難掌握，如果主辦單位沒有將會談論文贈送給圖書館，則讀者就很難取得這份文件，只有參加會議人員，才有此份資料。這類文件的檢索方法可比照上述第一項「文學圖書」。

(六) 網路資源：

網路上的文學資源越來越豐富，除了在作家個人網站上，可以尋獲相關文學資料外，有些雜誌社雖未製作光碟資料庫對外發行，但在自己的網站上公開部分雜誌內容，這時可利用網路搜索引擎，檢索到各類型文學資料，如《聯合文學》、《橄欖樹文學月刊》、《天下雜誌》等。網路上也可輕易的找到各種文學創作與評論的文獻，此即「網路文學」，這是近兩三年來蓬勃發展的新品種文學，讓文學新人有許多展示作品空間；有些網站為維持作品品質，投稿創

作需經過審核才可上網；有些作家一開始就不打算發行印刷式作品，直接將創作發表於網路上，供網友免費閱讀。網路文學已經是時代的潮流，越來越多的文學創作，將會直接在網路上公開。

七、文學史料數位化技術與服務的困難

文學史料數位化是一個頗具挑戰性的工作，它將面臨許多困難等待資訊技術去克服，且數位化服務觀念也是不斷地改進，因而服務人員必須自我提昇，才能應付資訊時代的潮流，底下分別簡述數位化工作可能需要面臨的困難：

(一) 系統建構：

甲. 作業平台：作業平台的選擇是一項困難的決定，因為大多數的文史資料工作者，並不清楚 Unix、Linux、NT 的優缺點，以及各家資料庫管理系統的差異性，在選擇作業平台時容易受到軟體廠商的左右，或者遷就預算多寡。基本上 Unix 與 Linux 作業系統相似，Linux 使用在中小型主機較多；NT 作業系統操作介面與 Windows98 較雷同，管理介面較親切，操作上較熟悉，軟體工程師比較容易尋獲，軟體開發費用相對較低。在資料儲存的安全性上，Unix、Linux 及 NT 作業系統可以使用磁碟陣列，保護資料的安全，當有一顆硬碟故障時，可以在不關機的狀態下，更換硬碟。在使用層面上，Unix 及 Linux





的主機比較能容許較多瞬間上網人潮，「乙」作業系統則對於使用人次有較多的限制。但無論如何，建議資料庫主機置放於機構之網路防火牆內，網頁服務主機則置於防火牆之外，較能保護主機的安全。

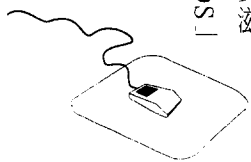
乙字碼問題：目前國內的網路使用者，其電腦大都使用 BIG5 碼，只能處理一萬三千四百六十一個中文字，再加上預留的造字區，頂多只能增加到一萬九千七百八十二個中文字，因此肯定無法滿足文字工作者的用字需求，單就文學家的名字當中就有許多 BIG5 碼無法顯示的字，如「極弦」、「琴涵」等，而文學史料經常需要處理日文或韓文的文獻，這時候如果主機內碼是 BIG5 碼者，就無法存入日韓文字，讀者的電腦自然無法顯示這些字碼，字碼不足問題一直困擾圖書資訊界，國內圖書館界自動化系統大都使用 CCCII 碼為內碼，可以處理五萬三千九百四十個字，雖然大部分的中文字都可輸入，並解決日韓文資料的建檔，但一般的使用者是以 BIG5 碼電腦連結圖書館自動化系統，因此也無法顯示畫面上的這些字。使用「國家中文標準交換碼（CNS11642）」，可以處理四萬八千零二十七個漢字，雖然已能處理大多數中文字，唯一遺憾的是一般讀者使用的電腦，並無該內碼，必須自行行政院主計處電子處理資料中心下載「CNS11643 中文標準交換碼全字庫」，才能使用或顯示該字集，目前這個字集使用較多單位的是戶政機關。近來資訊界

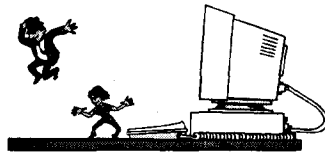
比較推薦使用 Unicode 當內碼，它是一個國際標準碼，可以容納七萬字以上的漢字，並且可以處理多國語文，且在國際上已經普遍使用，因此讀者的電腦只要使用 Windows 2000 以上作業系統（部分 Windows 98 作業系統也可使用），其電腦就可以輸入並顯示中文簡繁體字與日韓文字，雖然有些讀者的電腦不能鍵入 BIG5 碼以外的字，但是能接收主機端傳來的 Unicode 字碼，至少已能解決 BIG5 碼無法顯示中文字的問題。國內主要的入口網站或新聞網站如「PC Home」或「東森新聞網」等，已採用這種模式服務網友，只要在自己的電腦輸入「游錫」兩個字就可以找到行政院長的資料，這時如果網頁如能正確顯示顯示「游錫瑩」三個字者，大致上可以肯定該服務主機是採用 Unicode 內碼。

(二) 檢索系統：

甲 同義字：當我們在網路上查詢資料時，如果系統主機沒有建置同義字對照表時，使用者將會有些資料無法查到，例如「台與臺」、「豐與丰」、「莊與庄」等，最常見的例子是當我們鍵入「台灣」一詞查詢時，有些系統查不到以「台灣」著錄的資料，因此如果系統能自動處理同義字，可便利一般使用者。

乙 參照詞：參照詞表是針對自己開發的資料庫，建置一個可以相互指引的對照詞，例如將「愛滋病」、「愛死病」、「後天免疫不全症後群」、「AIDS」





設定為參照詞時，則不論讀者輸入那一個詞彙，系統都會去查詢所有的相關詞彙，這樣才能節省讀者之時間，又不會遺漏資料，但是系統維護人員經常必須維護這個參照詞庫，才能順利達成服務讀者的美意。

丙權威記錄：權威記錄檔是通常是在圖書館的圖書自動化系統，它可以針對個人作者、團體作者、題名、標題等款目進行權威控制，舉作家「小野」為例，所有他發表的作品，不論是用筆名「天牛」、「湯新」或者本名「李遠」等，都可以透過權威記錄的控制，將其所有的作品一次檢索出來，雖然讀者或許不知道「小野」還有其他的筆名，但是系統可以透過權威記錄檔，提供較完整的檢索服務，讓讀者可以查獲這位作家的全部作品。再舉一個團體作者的例子，例如「行政院文化建設委員會」「文化建設委員會」、「文建會」，甚至於將來如改制為「文化部」，都是指同一個單位，如果資料在建置時已經將其權威記錄關係著錄，則日後查詢資料時不但可將這一單位不同時期的資料一起檢索出來，如日後果真升格為「文化部」時，也不需要逐筆去修改原先已建檔的資料。

(三)系統管理：

甲著作權：數位文件著作權的問題，可能比其他資訊技術更難克服，尤其目前國內並無文字仲介團體可以負責著作權授權與著作權費用分配事宜，因此大多數的數位資料只能在圖書館內提供讀者服務，無法

公開於網路上，讓所有的民眾使用。這一兩年開始，出版社或雜誌社才開始注意著作權的問題，有許多雜誌在邀稿時就聲明，該文件投稿後的著作權屬雜誌社或出版社，或者要求作者同意讓該文件公開網站上，於是網路上的電子書及電子雜誌如雨後春筍般的大量出現，至於較早期的出版品，即便是原發行單位，也無法任意將書刊內容公開於網路上，這些早期的書刊，如果無法直接在網路取得，將使電子化服務的成效打折，因此盼望文字仲介團體能早日成立，才能解決授權問題，而將數位化文獻提供於網路上，當然屆時將會是有價的資訊服務。

乙電子媒體保存：由於資訊技術發展速度超乎想像，資料設備與儲存媒體都有一定之壽命，舉例來說早期的5 $\frac{1}{4}$ 軟式磁碟片，現在幾乎已經無法使用了，而當初如果有重要的資料存放在磁片上，應該儘早複製於新的媒體，如CD-ROM或DVD上；再者一般大型主機大概的壽命大概不會超過10年，個人電腦的壽命約5年，CD-ROM光碟片得壽命大約只有10年，不管是主機的更換或是儲存媒體的轉載，都是非常耗費人力的，尤其是主機的更換成本更高，如果主機的作業系統或中文內碼必須更換，更需耗費人力撰寫程式進行資料的轉出與轉入。前面提到資料備份的異地典藏觀念，也是近年來資訊界較認真去執行的觀念，以前各機構資訊單位通常將備份資料存放於機





房或資訊室，但這種作法不一定保險，萬一遇上水災、地震、火災或其他意外時，這些主要資料與備份資料統統毀於一旦時，試想不知該如何補救，如果這些資料與民眾的權益有關，該機構將不容易善了，且不知要虛耗多少人力、經費與時間，才能將資料重建起來，因此各機構應該要有電子媒體長久典藏的正確觀念。

以上僅列舉一些數位化時可能遭遇的困難，供同道參考，唯資訊技術發展迅速，說不定在未來三、五年內，這些障礙都已經排除，屆時從事文學史料數位化者可以更安心的使用數位化系統。

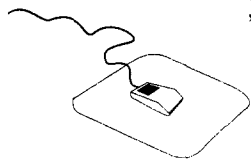
八、文學史料數位化的現況與發展趨勢

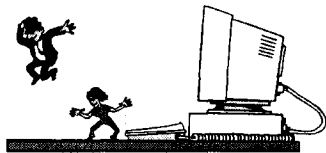
文學史料數位化在國內的發展，主要還是依賴政府的大力推動，例如國家圖書館建置之「當代文學史料影像全文系統」，以及文建會積極推動的「全國文化資料庫」，這些計畫都將會建置可觀的數位典藏成果，前者以優勢的圖書、期刊、報紙典藏品，將文學資料加以分析整理、掃描建檔，對每一位作家皆用同樣的尺度，竭盡所能的蒐集資料，那怕是小眾作家的相關資料，依然受到的公平待遇，是國內最具有代表性的文學史料網站之一。目前國家圖書館正進行作家資料調查，希望作家們協助該館就已建檔之資料，進行資料審查與增補工作，以強化該資料庫的時效性、

正確性與完整性，讓全民分享文學殿堂豐碩的果實。

後者正積極建置各類型地方文獻數位化工程，並統籌為「全國文化資料庫」建置計畫19，正式以國家級的數位典藏計畫，保存地方文獻與語文，各相關機構對於地方文獻蒐藏的資料類型，以照片最為普遍，其次分別為口述歷史、文物、方志、族譜、歌謠／諺語、私文書、報告書、詩文、公文書、碑碣、繪畫、見聞、海報傳單、書簡、戲劇、戰記、競選資料、日記、訃文、備忘錄等。「全國文化資料庫」開始執行之後，除將繼續民國八十八年的建置「地方文獻數位化計畫」的藏品，並擴增其藏品內容，所包括的資料類型有：音樂資料、美術資料、戲劇資料、舞蹈資料、傳統藝術資料、文學資料、電影資料、建築資料等八大類。其原始資料的來源有兩種：一為已經由中央與地方文化機關／單位收藏的公藏文化資料，二為尚藏於民間的文化資料。待「全國文化資料庫」逐年將各地方之文獻與文物建置後，文學工作者將有機會接近珍貴或未曾公開的文學史料。

至於「文建會」直屬機構「國立台灣文學館」，雖仍在籌備階段（預計民國九十二年中開館），但也積極扮演文學史料數位化工程，如《台灣文學年表》、《台灣文學作家作品集資料庫（作家全集）》、《台灣文學辭典》之編纂計畫，都在規劃建置當中，假以時日，必定有可觀的成果。





作家本人或其家屬在文學史料數位化領域的貢獻，也是不容忽視的角色，作家個人網站上可以提供不少文學相關資源，如向陽、張曼娟、金庸、吳淡如、彭懷真、廖玉蕙、琦君等之網站，相信未來還有更多的作家投入個人網站的建置，讓讀者有機會在網路上親近作家。

民間文學雜誌業者雖然在不景氣的經濟環境下，慘澹經營，但多少都有一份傻勁與執著，只要稍有財源，會毫不猶豫的投入數位化工程，如《文訊雜誌》、《聯合文學》、《橄欖樹文學月刊》等雜誌。

文學相關基金會與社團在文學史料數位化這個領域，也可以扮演重要的角色與影響力，例如「國家文化藝術基金會」、「蔣經國文化基金會」、「賴和基金會」等都長期支持文學相關活動，希望這些基金會能持續贊助文學相關活動。

在文學圖書業者方面，電子書的發展為雜誌出版業帶來一些生機，雖然國內網路行銷市場還有許多拓展空間，但總是一個契機，若非網路購物行為糾紛多，其實網路書店的確實有發展的機會。

網路文學的品質雖然還需加強，卻提供一個業餘文學創作者發表作品的園地，並開放網友討論網站上的作品，這是傳統文學出版品無法做到的，因此有人將這種文學稱為「網路新文體」，這是由於網路的特質而衍生的新文學品種，不管作者文學素養深度或文

筆洗鍊程度如何，它已經在網路上開創一片新天地，其中並不乏水平以上的作品，這種依賴網路生態的文學活動，將有繼續擴大的現象。

著作權是網路發展最大的障礙，也是文學傳播媒介最大的致命傷，但著作權卻同時保障文學創作者的權益，它像是一把刀子的兩刃，是一體的兩面，當我們保護著作權人的時候，同時也限制了知識的傳播。我們一方面要鼓勵文字工作者從事文學創作，並保障其智慧財產權，以獲取基本的酬勞，另一方面又為了要傳播知識，希望能將文學書刊，免費開放於網路上，供眾閱覽，這真是兩難。將來或許可以採用較折衷的方式，例如由文字中介團體在網路上代為收取使用費，再將其收益分配給作者或出版社，就能突破現有的限制，大大方方在網路上提供文學書刊的內容。

文學網路未來的服務趨勢應該是朝個人化、多元化、資源分享與付費制等方向演進。個人化網頁服務指的是個別化網頁設計，這個理念是為了讓每一位網路使用者，進到該網站後，感受到個別性、獨特性，網站設計者特別營造一個個人專屬服務之畫面，每一個會員或使用者，可以依照自己的喜好，設定畫面內容、偏好的主題、或設定直接進入指定的網頁，而系統主機也會主動與該使用者打招呼，例如多久沒有拜訪本網站或主動發問之提供該使用者偏好之訊息。在電子商務網站上的個人化網頁服務，更會紀錄會員





的消費偏好、消費能力等訊息，提供網站經營者寶貴的網路行銷的資訊，目前國內文學網站上尚未提供個人化服務，或許再過三、五年，讀者可以自行設定網站服務偏好，譬如有些讀者只偏愛新詩，系統就可以設計成每當該讀者上網時，就將畫面指定到新詩網頁或某一位詩人的網頁，享受到VIP的服務。

網路服務多元化指的是大眾網站與小眾網站都有其生存的空間，各自建置文學相關資源，各自服務不同類型之讀者，使文學網站多采多姿，造福廣大的網友。

資源分享是網路資源提供者最期盼的合作模式，只要大家都使用 Metadata 格式建檔，有朝一日，讀者可以在甲網站可以檢索到乙網站及丙網站的資源。

「知識是無價的」，未來的文學網路服務，可能有一部分文學資源是需要付費的，才能在網路上取得，為了讓文字工作者獲得基本酬勞，支付必要的著作權使用費，是可以接受的。當然讀者可以自行選擇網路上其他無償的資源或接受有償的資源。

九、結語

網路世界的最大特點就是「變」，每天都在改變，一切都在改變，誰也無法留住它，掌握它，唯有親近它、與它妥協或和平共存，才不會違背時代的潮流。

文學史料工作者通常對於資訊科技較為疏遠，這多少會影響數位化的推動腳步與發展成果，其實科技與人文是可以相處融洽的，雖然資訊科技可能帶來社會的失序、犯罪的溫床、價值觀的改變、倫理的喪失，但同時也帶動終身學習、遠距教學、知識普及、資源共享的優點，這道理與「水火」是人類的生存的必要物質，卻也是最大的災難來源相同，唯有善盡科技的優點，改善人類的生活品質，才是雙贏之道。

梁實秋先生曾說過：「文學發於心性、基於人性，亦止於人性。」（見《中華日報》副刊第十九版，民國八十五年四月七日），何懷碩先生也說過：「我們在文學藝術中看到人生的處境、人生的命運——看到我們自己，這就是共鳴。」（見《中華日報》副刊第十六版，民國八十八年六月二十日），可知文學是生活中不可或缺的一部份，作為文學工作者當善盡一己之力，致力於文學史料保存工作，才能讓後代子孫了解先民的文化活動與心路歷程。目前在各界的努力下，國內文學史料的數位化工作，雖然有一些小小的成果，但距離虛擬網路文學圖書館的境界，還有一些障礙要克服，且讓政府與民間通力合作，為國內文學活動注入新生命。

※此文承蒙台北科技大學中文系授權轉載，特此致謝！

