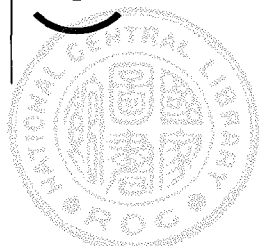




我國現代文學史料數位化典藏與服務(中)



莊健國·國家圖書館參考組編輯

四、文學史料數位化的特徵

當我們欲討論文學史料數位化的特徵時，應先明瞭傳統圖書報刊之典藏與服務困難：

- 一、分類編目耗費人力；
 - 二、典藏空間需求量大；
 - 三、資料不易長久保存；
 - 四、老舊書刊閱覽服務品質較差；
 - 五、早期書的書刊目錄索引比較不完整，資料查檢不容易；
 - 六、非書資料與特藏資料則更難服務讀者。
- 接著我們再進一步分析文學史料數位化有哪些特徵？是否可以解決傳統圖書館的服務限制：
- 一、數位化典藏服務軟體硬體投資成本極高：毫無疑問的數位典藏投入之人力與經費，是相當可觀的。
 - 二、數位化專業人力需求高：唯有專業人員的投

入與規劃，才能推出合格的數位典藏產品。

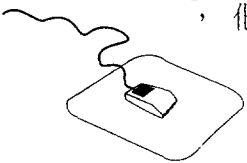
三、數位化資料複製容易：數位化資料最大的特點就是複製容易，攜帶方便。

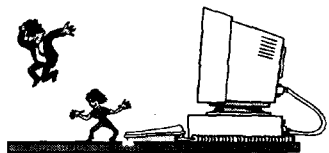
四、數位化資料查檢容易：數位化典藏最終目的就是提供迅速便捷的高品質讀者服務，因此資料檢索的便利一定是重要特色之一。

五、數位化資料節省典藏空間：數位化資料可以節省圖書館的實體空間，這是大家都肯定的一件事，尤其對於罕用之書刊，佔用圖書館太多空間對圖書館的發展，並不有利。

六、數位化資料不再侷限平面出版品，視聽資料亦可服務：除了圖書期刊等印刷式資料可以數位化之外，各類型多媒體資料也可以加以數位化後提供線上服務。

七、數位化資料典藏壽命較久遠：資料數位化後，因資料複製容易而可再轉製於新的儲存媒體上，





因此數位資源的長久保存是沒有問題的。

八、數位化資料可穿越空間障礙服務讀者：此即遠距圖書服務之概念，透過網路可以服務遠端的讀者，這是數位化後縮短服務距離的優點。

九、數位出版與網路行銷成為新興行業：數位化出版逐漸在網路新環境中形成，配合電子商務技術的開發，網路購書行為或閱讀行為已逐漸形成一股氣候，它已改變的傳統的消費行為。

十、數位化服務實現資源分享的理念：如果沒有著作權問題，數位服務早就實現資源分享的境界。

十一、數位化服務衍生網路著作權課題：為了保護創作者的智慧財產權，必要的限制似乎不可避免，因而網路著作權的議題，將會是網路世界中持續發燒的課題，如何在推動資訊公開與保護著作權人的權益當中取得平衡，還需考驗人類的智慧。

十二、數位化服務是未來的趨勢：無論如何，網路的時代已經來臨，數位服務是人類未來的發展趨勢，將來上圖書館、看電視、看電影、上班、開會等活動，都可以在家中的電腦實現。

五、文學史料數位化的規劃

當吾人開發一套文學史料數位化系統前，必須先思考下列課題：

一、該文學史料有無數位化的必要？

二、該文學史料能否數位化？

三、該文學史料如何數位化？

四、該文學史料值得數位化？

五、文學史料數位化的目的？

六、文學史料數位化的好處？

七、有無文學史料編輯專業人員？

八、有無持續維護數位化系統之資訊能力？

九、建置經費問題是否容易解決？

十、數位化持續建置與服務之能力？

在評估以上課題後，如果答案是肯定的，則可再進一步規劃下列需求：人力需求、技術需求、資金需求。

(一)在人力需求方面：

數位化過程必定會增加人力成本，包括人員培訓、系統觀摩、規格書制訂、軟硬體採購發包、書目資料建檔、影像資料掃描、影音資料建置、系統測試等。這是維持一個系統運作的必要人力需求，如果機構內尚無建置數位典藏的經驗，應儘早進行人才培訓，才能應付系統建置所需之專業知識。儘量避免以「急就章」或「撿現成品」方式規劃人力，以免將來系統完成後，還需面對現實的考驗，萬一人力資源無法掌控整個系統時，狀況輕者，會造成系統停頓，嚴重者，恐怕還需重新設計或調整作業，徒然浪費寶貴的時間與經費。文學史料數位化的人力基本需求有三





種，第一要有文學素養者，第二要有資訊素養者，第三要有行政能力者，前兩者負責規劃文學專屬系統，後者爭取預算與行政作業，三者缺一不可，相互支援與合作，才有可能在競爭的環境中順利完成任務。

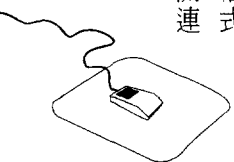
(二) 在技術需求方面：

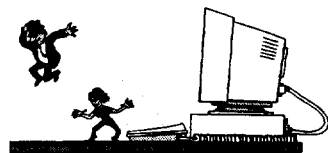
甲、軟硬體規劃：在技術需求方面，首先需規劃軟硬體相關設備，在硬體方面，需先計算文件數位化所需之儲存空間及使用人次，才能規劃主機等級及硬碟儲存空間大小，自大型主機到小型伺服器皆無不可，端看資料的重要性與使用率。系統軟體方面，如採用套裝軟體，價位較低，建置時程較短，唯系統功能一般僅能符合基本需求；或考慮開發專屬軟體，費用通常較高，開發時程較長，但較能符合建置單位之功能需求。另外還需依據作業平台，搭配資料庫管理系统 (Data base management system, DBMS)，大型主機大都採用 Oracle、Informix、Sybase 等資料庫管理系统，小型主機可使用 SQL 或 MYSQL 等，前者價位較高，後者價位較低，有些還是免費使用。

因此建置一個陽春型或小規模的系統，軟硬體初期投資大約只要 20 萬至 30 萬之間，全部開發時程只需三個月，就可以建構一個文學史料服務網站。而較具規模的系統通常需估算一、兩百萬才能規劃符合資訊安全與標準作業模組，包括採用雙主機服務，以防駭客入侵，將網頁服務主機置於機構防火牆之外，提供

一般讀者連線使用，而真正存放資料的主機置於機構防火牆之內，僅限機構內資訊人員及建檔人員使用，並以磁碟陣列規劃儲存之資料，以避免硬碟故障時，網頁服務必須暫停，另需規劃每週或每日備份建檔之資料，以防止主機故障時，可恢復先前備份之資料，而全文資料、影像資料、影音資料則應至少複製兩份以上於 DVD 或 CD 上，並保存於不同的建築物，即所謂之異地典藏，以防天災發生時（如地震、火災、竊盜），還有可能將另一個備份保存下來。

乙、著錄格式：關於文獻資料著錄格式，這幾年來在圖書館數位化系統中，大都採用詮釋資料格式 (Metadata)，詮釋資料是對文件或收藏品資料屬性的一種描述，目的在促進資料系統中對資料之檢索、管理與分析。這是考量跨機構、跨主機間之資料交換與整合，並考慮將來與國際接軌之需求，而採用符合國際標準之「都柏林核心集 (Dublin Core)」。這是一個簡單有彈性，且非圖書館專業人員也可輕易了解和使用的資料描述格式，它的著錄項目有 15 項，大致能滿足數位媒體資料編目要件，這 15 項包括：1. 主題和關鍵詞 (Subject)；2. 題名 (Title)；3. 著者 (Creator)；4. 簡述 (Description)；5. 出版者 (Publisher)；6. 其他參與者 (Contributors)；7. 出版日期 (Date)；8. 資源類型 (Type)；9. 資料格式 (Format)；10. 資源識別代號 (Identifier)；11. 關連





(Relation)；12. 來源 (Source)；13. 語言 (Language)；14. 涵蓋時空 (Coverage)；15. 版權規範 (Rights)。

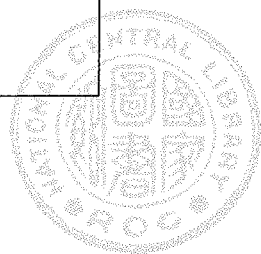
丙、數位化格式：

數位化典藏內容規格大致需考量下面四種資料格式（「地方文獻數位化之相關標準與作法」，《地方文獻數位化模式及相關標準研究報告》，南投市：台灣省文化處，民國八十八年。（<http://plisn.pnu.edu.tw/plis88/research/dl/5.html>））：

1. 文字資料 (TextualMaterials)：

文字資料數位化有兩種方式，一種是以影像檔格式儲存（如 TIF 格式、JPG 格式、GIF 格式），一種是以文字檔格式儲存（如 TXT、WORD、PDF、HTML 格式），前者製作成本低，且較節省時間，無須對文字進行校對，適合處理大量文件，但無法對資料內容進行全文檢索；後者製作成本較高與耗費時間較多，但是對於文件內容可以進行全文檢索功能，同時也較節省資料儲存空間。試分析比較此二者之差異：

項目	純文字格式 (txt)	影像格式 (tif)	備註
檔案大小	3~5KB	約 100KB (300dpi)	以 A 4 單頁 純文字為例
建檔時間	15 分鐘	1 分鐘以內	約 1500 字
建檔費用	較高	較低	
全文檢索	有	無	
文字校對	要	無	
文字排版	要	無	
文字美編	可	無	
造字問題	有	無	
影像品管	無	要	
資料內容	容易擅改	不易變更	
其他限制	無法處理特殊符 號或圖像資料	檔案較大不便做 成電子書	





經由上述之分析比較，可知資料內容如不涉及全文檢索時，採影像掃描方式儲存資料較節省時間、人力與經費，例如國家圖書館之「當代文學史料影像全文系統」、「中華民國期刊論文索引」及「傳記文學」光碟資料庫即採此一模式建檔；而當文件內容需要做提供檢索之功能時，唯有採用純文字格式建檔方能解決，例如中央研究院之「二十五史資料庫」、香港迪志文化公司之「四庫全書電子版」之文字建檔格式。

PDF (Portable Document Format) 檔案是近年來另一種熱門儲存文字格式，它保有純文字格式全文檢索之優點，又有保護文件內容不被修改的功能，甚至有限制列印及存檔的功能，國家圖書館「全國博碩士論文系統」收錄之論文電子檔即採用 PDF 格式製作。

2. 圖像資料 (Pictorial Materials) :

常見的相片、圖片、手稿、畫冊、海報等資料數位化的主要格式，有下列幾種，一般都是採用掃描機來製作：

- (1) TIFF：適用黑白、灰階、全彩的文件資料，壓縮效果佳，適合處理舊籍或檔案資料，並可儲存為保存用圖。
- (2) GIF：適用於黑白、灰階、256 彩圖，壓縮效果亦佳，適合處理預覽型圖片。
- (3) JPEG：適用於灰階與全彩，壓縮效果極佳，適合提供於網路上公眾瀏覽。

(4) BMP：適用黑白、灰階、全彩的文件資料，壓縮效果普通，適合儲存為保存用圖。

(5) PDF：適用黑白、灰階、全彩的文件資料，壓縮效果佳，適合處理整份文件、整本圖書或期刊，並方便加入浮水印，防止資料被冒用或擅改，主要提供於網路上的電子書。

目前在國內外，對於文件的保存方式，大都參酌美國國會圖書館的作法，將一份圖像同時儲存成三種格式：

(1) 預視圖 (Thumbnail image)：以 GIF 格式儲存，供網路上預覽。

(2) 檢視圖 (Serviceimage)：以 JPEG 格式儲存，供網路上瀏覽。

(3) 保存用圖 (Archivalimage)：以 TIFF 格式儲存，作為永久保存用。

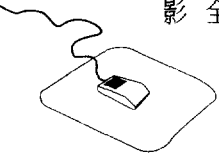
3. 聲音資料 (Soundrecordings) :

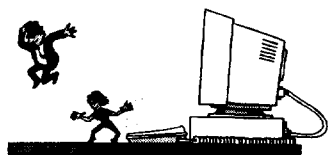
聲音資料採用 WAVE 格式、MP3 格式、RA 格式居大宗：

(1) WAVE：檔案較大，一分鐘的聲音需佔用 10 MB 的空間，適合用來保存原音。

(2) MP3：最大優點是檔案小，唯略有某種程度失真。

(3) RA：最大優點為線上播放，不用等到檔案全部下載完畢，最大缺點是容易受到網路品質好壞而影響。





響收聽品質。

4. 視訊資料 (Audio-Video Materials)

視訊資料的處理需要之特殊設備與專業人員，處理物件拍攝、影像編輯與剪接之能力，目前主要的格式有：

(1) AVI：適合處理畫質要求較低階之產品。

(2) QUICK-TIME：適合處理畫質要求較低階之產品。

(3) MPEG：這是最普遍的視訊格式，目前我們最常使用之VCD片中的影音格式，就是採用這種格式，它提供極高的壓縮比例與較高的畫質。

(4) Streaming Video：這是可透過網路上直播的一種格式，也是容易受到網路品質好壞而影響收訊品質。

(三) 在資金需求方面：

建置一套數位化系統，除了上述人力需求與技術需求外，經費的來源也是非常重要的，這三者其實缺一不可。資金方面的取得，以政府單位為例，大多經過需先提出計畫書，再透過預算之編列、審查與核定後，才能正式執行計畫，這中間約需兩年以上的作業程序，等到系統正式上線完成後，應該才是困難的開始，這時候如何再去申請經費，維持系統正常運作，也是一門學問，或可考慮再以相同名義申請第二期發展計畫，但得看上級單位支不支持該項計畫，否則就



變戲法的哲學家

對於康德從摧毀宗教又回到上帝和靈魂不死的宗教觀點這種矛盾，尼采的一位友人保爾·瑞評論說：「在閱讀康德著作的過程中，你會產生置身於鄉村集市的感覺。在他那兒，你能買到所需的一切——意志的自由與束縛，唯心主義與對唯心主義的駁斥，無神論與善良的上帝。就像魔術師用空帽子變戲法一樣……」

只有想辦法自籌經費，勉力維持系統運作。這也就是我們在網路上經常看到一些網站，網頁內容無法定期更新的主要原因。其他可以考慮的運作方式，譬如以產品發行來自籌經費、尋找廠商合作建置、委託開發、尋求金主（基金會）、商品化行銷等方式皆可，端看其數位化產品有無市場經濟價值，例如故宮博物院的文物複製品與光碟產品，在市場上反應較熱烈，就不難尋求合作開發的對象，而文學史料就沒有這麼樂觀了，除了金庸的小說，或少數熱門作品外，一般都很難有創收，筆者日前在一家量販店無意中發現整套文學電視劇「人間四月天」10卷錄影帶，只賣一九九元（原價二九九〇元），心中感觸良多，文學工作者的酬勞，應該是不大穩定的。

□

