

詮釋資料與數位化古籍的著錄

顧力仁

國家圖書館 編審

摘要

國家圖書館在完整的善本書目紀錄以及多年投入文獻影像處理所獲致的經驗為基礎上，建置出兼含詮釋資料 (Metadata)、古籍影像以及全文資料庫在內的數位資源檢索系統。國家圖書館所建置的古籍影像系統，在書目尋檢上，係採用 Metadata 格式。整理古籍 Metadata 的內涵，並試圖建立標準化的共識，對存世古籍聯合目錄的建置有密切的關係，而古籍聯合目錄的建置正是吾輩關心文化遺產、關心華文資源共享、關心網路數位知識工程的圖書館員所應承擔的歷史使命與現實責任！

關鍵詞：數位化古籍，Metadata，詮釋資料，後設資料

一、珍善古籍數位典藏與詮釋資料

文獻資源的數位化以及網路技術的進步，不但改寫了圖書館的服務型態，影響所及，也使得古籍的整理、組織及利用起了革命性的變化。加諸現代公民重視歷史記憶及文化資源，使得圖書館所珍藏的歷史文獻重新躍登網路而為重要的利用素材。有鑒於網路中文資源亟待充實，台灣地區於 2000 年起針對重要文化資產單位所保存的珍貴文物規劃「國家重要典藏文物數位計劃」，至今已有五年，政府對此殊為重視，斥資台幣上億，希望結合典藏機構、研究單位以及產業界多方資源，豐富網際網路的本土內容，對內以保存文物影像、促使文物資訊加值，進而提昇全民精神生活為目標，對外則希望充實網路文化內涵，平衡並傳播中華文化。¹

國際間在上一世紀末即注意到保存人類文化遺產的重要性，陸續推動若干大

¹ 見《數位典藏國家型科技計畫》計畫緣起，http://www.ndap.org.tw/1_intro/history.php

型的國際及國家型計畫，例如：聯合國「世界文化遺產數位化計畫（UNESCO Memory of the World）」、美國國會圖書館「國家數位化圖書館計畫（The National Digital Library Program）」；此外，各國尚有多項針對珍本古籍和手稿等珍罕文獻進行的數位化保存計畫，例如：大英圖書館「十一世紀史詩手稿數位計畫（The Electronic Beowulf）」、IBM 公司「梵蒂岡圖書館計畫（The Vatican Library Project）」，……。² 在國家科學發展委員會的推動下，台灣地區的國家圖書館和其他兩所同質機構，分別是中央研究院及故宮博物院，啟動了本地區珍善古籍的典藏數位化工作，並且協同建立合作機制，共謀本地區重要善本古籍的數位存護工程。³

中央研究院歷史語言研究所傅斯年圖書館、故宮博物院圖書文獻處以及國家圖書館特藏組早於五年前即協商妥訂以「明人詩文集」以及「古籍附圖」作為三個典藏機構珍善古籍數位化的重點，希冀經由群體合作，建立台灣地區重要漢學數位資源的特色，並提供海內外研究所需⁴。經過多年的努力，傅斯年圖書館開發出若干實用的珍善古籍數位系統，並且成功的建立了一套自實體管理到數位典藏，乃至於虛擬檢索的作業流程；而故宮博物院也建置完成以「古籍附圖」為主要內容的善本古籍影像資料庫；國家圖書館則在完整的善本書目紀錄以及多年投入文獻影像處理所獲致的經驗為基礎上，建置出兼含詮釋資料（Metadata）、古籍影像以及全文資料庫在內的數位資源檢索系統。⁵

國家圖書館推動「善本古籍典藏數位化計畫」已有五年，數位產出(截至 2004 年 10 月止)簡介如下：⁶

1. 詮釋資料（Metadata）26,619 筆，包含叢書子目的書目 13,000 筆在內，該詮釋資料是以 Dublin Core 十五個基本欄位為架構，但為了將善本古籍豐富的屬性精確地描述出來，在相關的欄位下，同時搭配欄位修飾語（element qualifier）。（見圖 1）

² 談及台灣地區和大陸地區的古籍資源數位化可以參閱以下二文：

(1)鄭賢蘭、薛文輝，〈古籍、特藏書目數據庫及數字化調查統計表〉，《文津流觴》第八期。

(2)鄭恆雄、吳敏萱，〈台灣地區古籍文史電子參考資源資訊系統建置之分析〉，載《2004 年古籍學術研討會》(台北市：輔仁大學，民 93 年)，會議論文。

³ 上述三個單位組成善本古籍主題小組，為數位典藏國家型科技計畫內的內容發展分項計畫主題的十二個小組之一。

⁴ 見《數位典藏國家型科技計畫》內容發展分項計畫的〈善本古籍主題典藏〉，http://www.ndap.org.tw/2_catalog/visit_folder.php?id=386

⁵ 上述三個系統詳見：

(1)宋慧芹，〈國家圖書館古籍影像檢索系統概要〉，頁 17-28。

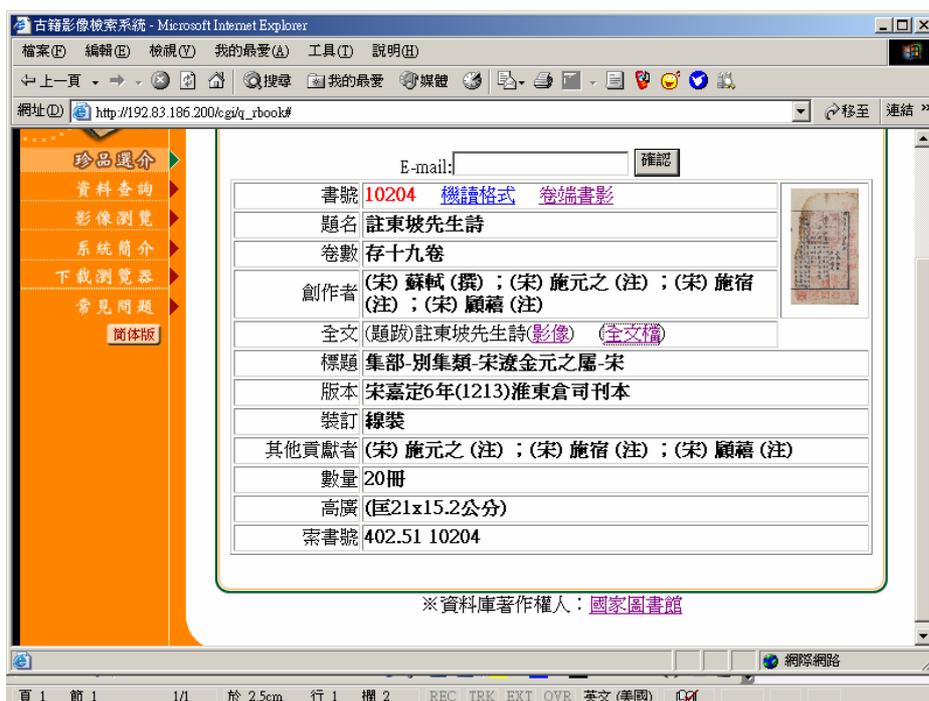
(2)林妙樺，〈中研院史語所傅斯年圖書館數位典藏系統之發展現況〉，頁 29-36。

(3)吳壁雍，〈故宮善本古籍影像與後設資料庫系統建置現況〉，頁 43-48。

載於國家圖書館編，《古籍聯合目錄資料庫合作建置專集》(臺北市：編者，民 92)。

⁶ 參考宋慧芹，〈國家圖書館古籍影像檢索系統概要〉，見註 16 的 (1)。

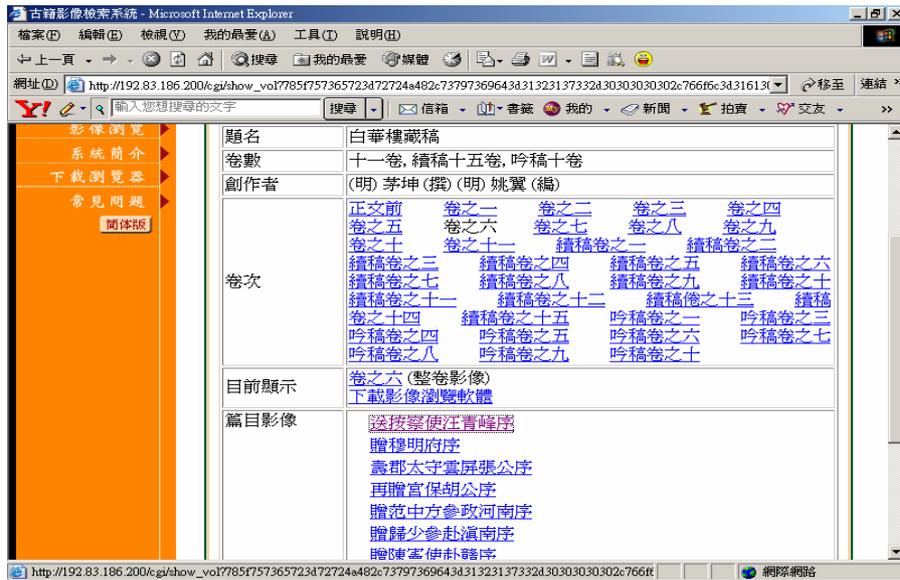
2. 數位影像 274 萬影幅，包含彩色（72~150dpi 及 300dpi）4 萬影幅以及黑（300dpi）270 萬影幅，館藏善本 8,338 部已完成數位影像，佔全部善本的 69 %。
3. 全文資料庫 10,936,583 字，包含篇目全文 2,097,545 字以及題跋全文 64,991 字。館藏善本多有篇目，每部的篇目多寡不一；另館藏善本富含明清藏家及學者的題跋，題跋全文有標點，可助辨識，而利研究。



圖一：國家圖書館推動「善本古籍典藏數位化計劃」檢索畫面

國家圖書館所建置的古籍影像系統，在書目尋檢上，係採用 Metadata 格式，同時可點選顯示機讀書目格式，並且在網路上展現每部善本的卷端書影。從文獻內容揭示的層次來看，第一層次以 Metadata 格式顯示書目的基本且為重要的特徵，第二層次顯示各該善本的卷目，第三層次顯示該卷內所有的篇目，最後顯示文獻的影像（可選擇單看一篇或該卷各篇）。上述書目查閱子系統在架構及介面的設計不單單考量到一般讀者的需要，也兼顧到圖書館古籍管理館員的需要，讀者可以循序而進的看到一部古籍的書目、卷次、篇目以及影像全文；（見圖 2）而館員可以從 Metadata 及機讀格式、卷端書影所提供的信息進行書目抄錄作業或參考比對。⁷

⁷ 近知海內外有許多收藏古籍的圖書館，經常上網查詢國家圖書館所建置的館藏古籍影像系統



圖二：卷之六的各篇目，在此可點選任一篇目或者其他卷名

二、數位化古籍的著錄

數位化古籍的書目著錄有兩個重點，一個是書目格式，一個是書目品質。國家圖書館上述善本數位化典藏系統的書目採用 Metadata 格式，以利網路資源傳佈。此格式係以國際通用的 Dublin Core 15 個欄位為據，並將國家圖書館原已鍵好的機讀格式書目檔和 Metadata 格式訂定欄位對照表，再由機讀格式轉成 Metadata 格式，由於善本書目內容複雜，其有 15 個欄位不盡容納者，就性質所近併入相關欄位內。國家圖書館所藏善本早於 70 年代即建妥機讀格式書目檔，⁸ 復經補充整理，雖與盡善的目標仍有距離，但書目內容的豐富與精詳頗受其他同質友館的肯定。

善本古籍的書目品質有賴「著錄」與「考訂」兩個因素，古籍著錄與一般圖書無異，重在正確揭示出文獻的特徵，提供使用者對古籍的內容作進一步的整理、開發與利用。正確的著錄當先有標準化的規範，不同地區使用的規範包括台灣地區的《中國編目規則》善本部分、《中國機讀編目格式》；大陸地區的《中國文獻編目規則》古籍部份（正在修訂中）、《漢語文古籍機讀目錄格式使用手冊》、

以及中文古籍書目資料庫這兩個資料庫。有北美某著名大學東亞館的館長告知筆者北美多館進行館藏古籍編目最常使用的資料庫即為中文古籍書目資料庫，另英國也有專治古籍書目學者來信表示對上述資料庫的內容頗重視。

⁸ 國家圖書館曾參與善本圖書機讀編目格的訂定，詳見中國機讀編目格式工作小組，〈中國善本圖書機讀編目格式一實驗報告〉，《國立中央圖書館館刊新 15 卷 1、2 期》（71 年 12 月），頁 28-35。

《北京大學古籍數字圖書館古籍元數據標準著錄方法》以及北美地區《RLG 中文善本書國際聯合目錄項目編目規則（草案）》。即便有了標準化的著錄規範也不能一定確保書目品質，這個說法對善本古籍尤其明顯，所以古籍考訂在書目品管中扮演很重要的角色，舉「版本項」為例最能道明「考訂」的關鍵性。

Metadata 是因應網路資源檢索與共享的書目資訊表達方式之一，與機讀書目格式一樣，要以文獻本身的著錄要項，也就是應該要揭示出來的文獻特徵，作為書目資訊的來源。我們縱有機讀格式，但不能取代編目規則；同樣的，我們研創 Metadata，但還是保有編目規則。據聞北京圖書館刻在修訂《中國文獻編目規則》的古籍部份，誠屬必要，應予重視。⁹

目前古籍 Metadata 呈現的著錄方法，包括中央研究院傅斯年圖書館的「善本書詮釋資料及著錄範例」以及北京大學「古籍數字圖書館古籍元數據標準著錄方式」兩種，國家圖書館則採取機讀格式轉成 Metadata，並訂有善本古籍 Metadata 與 CMARC 欄位對照表，海內外其他友館也正在密切注意並積極訂定相關著錄標準。若是從合作的角度來看，現在未嘗不是一個時機，是一個群策群力探討中國古籍文獻基礎著錄及其適應網路古籍數位資源易於尋檢共享的好時機，中國大陸、北美以及台灣地區都有許多學理的探討和務實的做法，特別是近十數年來古籍的考編工作受到前所未有的重視，以紙本或電子載體呈現的重要古籍書藏不下十餘種，例如：中國國家圖書館古籍書目數據庫（2001-2004）、國家圖書館善本書志初稿（1996-2000）、香港中文大學圖書館古籍善本書錄（1999）、美國哈佛大學哈佛燕京圖書館中文善本書志（1999）……。至於其他以「聯合目錄」形式呈現的，例如：中文古籍書目資料庫（1998-）、香港所藏古籍書目（2004）、東北地區古籍線裝書聯合目錄（2004）……，所在多有，此不贅敘。值得我們關注的，不僅是這麼多體例精純、考訂詳確、內容閱富的古籍目錄，而是各目錄內的書目紀錄「詳、簡不一」，而這些數以萬計「詳、簡不一」的古籍書目，當如何把他們歸入到 Metadata 的相關欄位去，使得書目資訊的本身各得其所，也使得尋檢這些古籍書目資訊的使用者能各取所需？

核心書目（或稱核心元素）是解決的方法之一，這個問題的解決不祇是處理古籍 Metadata 格式的內涵，並試圖建立標準化的共識，它還對存世古籍聯合目錄的建置有密切的關係，而後者，不也正是吾輩關心文化遺產、關心華文資源共享、關心網路數位知識工程的圖書館員所應承擔的歷史使命與現實責任！

⁹ 中西方保存珍本古籍所反映出來的文獻信息有異曲同工之效，據報導知拉托維亞科學圖書館進行古籍數位化時，其電子目錄可顯示文獻的裝幀、印刷形式、藏章、殘缺、墨迹、批注、插圖繪圖、作者姓名第一個字母的寫法、書頭書尾、章頭章尾的小花飾。見段潔濱，〈拉托維亞科學圖書館古籍手稿保護數據庫一體化〉，《江蘇圖書館學報》（2001.06），頁 54。