

技術變革的焦慮與跨越—從宋代印刷術的過渡性衝擊談當代古籍文獻資訊化的大數據未來

The Anxiety and Transcendence of Technological Revolution: Discussion on the Big Data Challenge of the Modern Digitization of Chinese Ancient Documents from the Transitional Impact of Printing Technology in the Song Dynasty

劉寧慧 Liu Ning-hui

國立臺北大學古典文獻與民俗藝術研究所副教授

Associate Professor, Graduate Institute of Chinese Documentation and Folk Arts

Email: mimiliu@mail.ntpu.edu.tw

【摘要 Abstract】

古籍文獻數位化從 1984 年臺灣中央研究院「二十五史食貨志」的製作到今天，已歷整整 30 年。30 年來文史學界投入嘗試與製作，追隨電腦科技的發展腳步，盡力與盡情將原屬紙本及古典形式的古代典籍轉換成數位訊息、仿真圖片以及自動檢索的文本資料。大家都明白以及相信，數位化必然是下個階段的古籍整理方式與里程碑，30 年來各種時新製作也確認了這個趨勢。不過，相對其他領域的革命，古籍文獻走的步伐還是保守、有限，主要還是在初階的廣泛全文化及圖版掃描，文本內在的關連精神與外在的形式跨越都尚屬稚嫩。而從較早不少個人化的建置到現今，漸漸形成資源單位的專利以及商業集團的獨大，文史專家投入逐漸減少，遑論整個理應迎向數位化到來並主導古籍文獻過渡的文史學界。究其原因，不外發現了古籍數位化工程的龐大複雜、資訊技術的難以掌握以及個人力量的侷限。這種趨勢與致力相違的情況，更深層的原因是沒有真正看到數位化革命的效益與新天地，而闕於認識的緣由又肇因於對技術的陌生。本文採取重回唐宋印刷術發展的歷史觀察策略，從當時所面臨的媒介革命、遭遇的實務問題以及改變焦慮等，看今天我們所面對的類似景況—技術的強大與強迫、從事的艱辛與

低效、變化的疑慮等。更有甚者，如今的數位革命已跨步向資料革命，又是一個人類的新經驗在形成中。古籍文獻作為一種鉅量資料，在資訊技術的運算下將呈現何種新認識？它們的意義在哪裡？希望藉由歷史的發展與鼓舞，找到今天努力跨越的勇氣與信心，促進文史學界往新時代大步邁進、夢想齊發。

Since the production of “二十五史食貨志” of Academia Sinica in Taiwan in 1984, to date, digitization of Chinese ancient documents has been developed for 30 years. Over this period, the literature and history field has attempted to use computer technology to transform Chinese classic ancient documents into digital information, replicated images, and texts of automatic indexing. It is obvious that digitization is the future trend and milestone of reorganization of ancient documents. Over the past three decades, various new production methods have supported this trend. However, comparison to revolutions in other fields, the progress of Chinese ancient documents is rather conservative and limited. It is mainly based on primary full texts and scanning of images. The essence of texts and external forms are still at an early stage. From early personalized construction to the present, the focus has been shifted to patents of resource units and oligopoly of business groups, with less participation of historic experts, not to mention the academia. The reason is that ancient documents digitization is complicated and vast, thus, the complexity of the information technology and limitation of individual efforts have led to the inferior result. Moreover, the outcome and new opportunities of digitization revolution are neglected due to the unfamiliarity with the technology. This paper observes the historic development of printing technology in Tang and Song Dynasty. It further compares media revolution, obstacles in practice, and anxiety of change at the time with current similar situation: power and compelling force of technology, difficulty and low effectiveness, anxiety of change, etc. In addition, modern digital revolution is becoming data revolution. It is the formation of human beings' new experience. What can be the new topical subject in Chinese ancient documents, as Big data, in calculation of information technology? What is the significance? By historic development and encouragement, this paper intends to encourage the literature and history field to approach a new era.

【關鍵詞 Keywords】

印刷術、版刻、雕板印刷、古籍、文獻、數位化、資訊化、技術革命、大數據
Typography; Chinese Ancient Documents; Digitization of Chinese Ancient Documents; Woodblock Printing; Big Data; Technological Revolution

一、前言

人類實在是聰明無比的動物。在他的文明歷程中，人類不斷發現他在生活上的問題並且一再研發出改善的事物與方法，既改變了當下的生活型態，也推進了整體文明的進步。而我們相信，這個過程中所執行過的嘗試與實際發展，絕不僅我們今天知道、看到且受益的文明內容，它的經歷既非文化史中敘述的那樣流暢，它的選擇也不是一開始就眉目清晰、理所當然，其中曾存在著許多疑慮、困惑與掙扎，一路上是摸著石子過河。最後，歷史做出抉擇，有些只如流星一瞬劃過天際，有些在價值選擇中勝出，成為人類一步步墊高文明的基石，至今為我們所喝采。

今天我們又立在文明的十字路口，在一波波人類空前的文化巨浪前。電腦的出現與快速進化，不過半個多世紀，就把幾百年來習慣的生活與世界劃開一道深溝—使用電腦與不使用電腦兩種人口。智慧型手機與社群網站的出現，再把這已然分割的板塊推移到兩者難相遙望。數位世界那一頭，日新月異，雖然發展的腳步還沒緩止，但我們知道它就是朝任意夢想飛馳而去，熱烈暢旺。另一頭，則是面積迅速在消蝕，闐暗寂聲。我們又一次面臨判斷以及行動採取，是迎向還是固守？是大破大立還是微調？

古籍文獻數位化的製作，最早是 1984 年臺灣中央研究院「二十五史食貨志」計畫。彼時迄今三十年，文史學界有所嘗試與建置，跟隨電腦技術的腳步，階段性地將原屬紙本以及古典形式的古代典籍轉換成數位訊息、掃描圖片以及自動檢索的電子資料庫。大家都明白以及相信，數位化必然是未來古籍整理現代化的主流與里程碑，三十年來的製作也確認了這個趨勢。不過，相對其他領域的革命，古籍文獻走的步伐還是顯得保守與有限，主要層面未脫初階媒材轉換，如圖檔掃描、全文輸入，後設資料的詮釋也是各異的。這些初步的製作，對於文本間的串連、格式的創新與全文內在關

連等，深度的數位建構並不足。¹ 數位化已走進本質性地改變——即對閱讀方式、思考方式、研究方法與相應連帶系列課題的顛覆。從過去的遙望與預見，今天它根本已登堂入室。

作為材料管理的主人，文史學界進行了何種因應？較早還有不少個人化的建置，之後則逐漸趨向資源單位、商業集團與公益開放社群的鼎足，文史專家在其中的身影有限，遑論整個理應迎向數位化到來並主導古籍文獻過渡的文史學界。臺灣的情況尤然，古籍數位化發展幾呈停滯狀態，近年甚至見不到多少相關論文發表。² 中國大陸的數據庫則成長快速，商業產品推出豐富，有關的探討也愈見深度。不過，整體來說，仍存在相當侷限，包括所成就的數據品質以及所發揮的影響性。究其原因，不外發現了古籍數位化工程的龐大複雜、資訊技術的難以掌握以及少數人力量的未足。文史界這種趨勢認知與致力相違的情況，更深層的原因是沒有真正看到數位化革命的效益與新天地，而關於認識的緣由又肇因對技術的陌生。

類似的境遇在歷史上並非首度面臨。上一個文獻革命——印刷術的發明，就形成過同樣精神的焦慮與猶疑，只是程度不同。如果我們對今天的開展有所徬徨，或許將時間視野拉大，反思歷史的命運，也許可以看到曙光並生發邁步的信念。本文不打算從媒介改變來討論整體數位化的文化發展，而是聚焦在技術變化所帶來的刺激與挑戰。

³ 策略上採取重回宋代印刷術發展的歷史觀察，從當時所面臨的媒材革命、遭遇的實

¹ 古籍文獻領域的製作，近三十年成就主要有 1999 年正式啟動並持續十餘年之「數位典藏與數位學習國家型科技計畫」，其中如「考古」、「檔案」、「拓片」、「善本古籍」、「漢籍全文」等主題，都有不少單位進行數位化建置。另有「數位加值資源整合」主題，亦不乏以上數位化成果之加值利用。它們的內容主要是文物、文獻的掃描與後設資料詮釋且以典藏品為主，不是大規模與廣範圍的古籍文獻數位整理。又如文化部「國家文化資料庫」及國家圖書館「臺灣記憶」等網站，整合、蒐羅與展現的也是臺灣相關人事時地物典藏資料，以圖片、影音、後設詮釋為主，部分有地理資訊系統的結合呈現，但都未及文本內容語意的細緻分析與連結。

² 「數位典藏與數位學習國家型科技計畫」，幾個存藏古籍的公藏單位如故宮博物院、傅斯年圖書館等都展開了善本的掃描以及後設資料建置計畫，但此國家經費補助已停止。目前所見數位典藏網站公告顯示，「『數位典藏與數位學習國家型科技計畫』已於 2012 年年底執行完畢，自 2013 年 1 月 1 日起，本網站內容將不再更新訊息。」

³ 近數十年古籍文獻數位化成果，可參見筆者與編之「國際漢學研究數位資源選介」（臺北市：漢學研究中心，2010）。

務問題以及改變焦慮等，看今天我們所面對的，事實上是更為艱鉅的景況——技術的強大與強迫、從事的艱辛與低效、變化的疑慮等。⁴

如今的數位革命已歷經幾番進化，它的本身即是驚奇。比如近期提出的「大數據」概念，宣告數位革命跨向資料革命，又是一個人類的新經驗在形成中。本文以一個古籍文獻整理與資料庫參與製作者的立場，提出現階段古籍數位化技術與觀念突破的粗淺觀察與省思，目標不是連專業資訊人員都難以即時掌握的各類技術，而是製作上來自文史資料端的震撼與期待。古籍文獻作為一種鉅量資料，在資訊技術的連結下可以出現何種結構的文本？在秒速運算下將展現哪些新認識？如此迎向、追求的意義在哪裡？希望藉由歷史的發展與鼓舞、技術船堅砲利的支撐，找到今天努力跨越障礙的勇氣與信心，文史學界一樣可以往資訊新時代大步邁進、夢想齊發。

二、宋代雕板印刷術的「庸俗化」傾向

雕板印刷術被視為古中國四大發明之一，並由此改變了全人類歷史，締造了世界新文明。它的發明是令人鼓舞的，作用是教人讚嘆的。不過，這是後世觀點。在它同時的人們，事實上經歷了一個從驚奇、疑慮到嘗試並最終適應的過程，它不是一開始就十分篤定的，掌聲也曾稀落，譏評之聲不絕於耳，猶如人類文明史上所經驗的任何新事物。在雕板印刷術應用於圖書製作前，所有典籍文獻都是透過手寫完成，今天看來，技術改變是偉大發現，工具便利是王道，但從北宋人的眼光來看，這卻曾是一場傳統崩壞、儒雅傾頹、流俗張揚的「災難」。⁵

⁴ 類似理念與方法的寫作，還有張靜、王兆鵬。張靜、王兆鵬（2010）與本文的大方向及表現策略接近，同樣選擇從歷史上媒材革命經驗看今天的境遇，該文以北宋初百年間情況為主，重點在文人與朝廷對雕版印刷新事物的接受及態度，認為北宋初已形成全媒體出版概念，然雕版印刷終將獲得主流地位並大力推動宋代文化轉型。再者類比當今數字化潮流，提出數字化出版將再度形成社會文明推動力，新文化即將到來。而本文重心則在技術層面造成的影響，討論數位文本型態趨勢與新文獻學，未及出版文化。

⁵ 早在 2009 年，吳夏平即提出「最初的刻本書籍同樣是被視為俗的，不為文士所重」（吳夏平，2009）。

(一) 唐宋儒雅的手書傳統

雕板印刷術開展於唐代初期，但真正較為普及起碼要到兩百多年後的北宋後期。在北宋之前，自漢以來，圖書、作品的流傳與複製主要依賴手寫傳抄，另則是刻石。當時的知識與學問侷限在少數知識份子手中，文字能力與學術有大範圍重疊，即使是唐代，情況仍大抵如此。⁶ 六朝與隋唐的士人，擁有典籍的途徑經常是靠「手自鈔錄」。抄書以獲書的目的還在其次，最重要的，這是傳統學問的方法，所謂「手鈔口誦」。最知名的例子，南朝梁王筠，一生中長達 46 年的圖書抄錄經歷。他說自己：

幼年讀五經，皆七八十遍。愛《左氏春秋》，吟諷常為口實，廣略去取，凡三過五抄。餘經及《周官》、《儀禮》、《國語》、《爾雅》、《山海經》、《本草》並再抄。子史諸集皆一遍。未嘗倩人假手，並躬自抄錄，大小百餘卷。不足傳之好事，蓋以備遺忘而已。(姚思廉等，1983，卷 33，頁 486)

王筠從少到老，喜歡書籍，經典讀七、八十遍，特愛的《左傳》撮鈔五遍，重要作品也有兩遍、一遍的記錄，他對這樣精勤的用功謙稱只是備忘而已。南朝齊宗室蕭鈞，儘管朝廷中富有典籍，但他還是經常「手自細書」，寫成巾箱小本，一則攜帶與檢閱方便，另則「一更手寫，則永不忘」(李延壽，1985，卷 41，頁 1038)。六朝這樣抄讀成學的例子不少，見諸記載的，如晉葛洪「夜輒寫書誦習，遂以儒學知名」(房玄齡等 a，1983，卷 72，頁 1911)，晉代范汪為人傭書，「寫畢，誦讀亦遍，遂博學多通」(房玄齡等 b，1983，卷 75，頁 1982)，南朝梁王僧孺「寫畢，諷誦亦了」(李延壽，1985，卷 59，頁 1460)。唐代也承傳這樣的方法，張參「為國子司業，手寫九經，每言讀書不如寫書」(王瑞來點校，1983，卷之一，頁 3)。即便貴如萬乘之尊的天子，唐高宗李治也曾「親洒宸翰，遍寫《九經》」，因為「讀十遍不如寫一遍」(王瑞來點校，1983，卷之一，頁 3)。

再者，六朝隋唐皆講究書法，不管是書工傭書還是文人親寫，對書法的留意不乏記錄，事實上在書寫時代此為基本素養。⁷ 特出的如唐柳公權姪兒柳仲郢，公餘布卷

⁶ 在此，出現「傭書」書工的問題。我們今天所了解的六朝書工、書手，多來自史書記載，這些能載入史籍的人物絕非泛泛，為人傭書，經常是早年家貧之故。唐代民間書工發展更為商業化，的確不乏純為利益的筆工，本文乃以多數士人階級情況而論。

⁷ 在孔凡禮點校(2005)之「唐書判」中談及唐代銓選擇人之法有四，其中之一即「楷法遒美」，並結論道：「既以書為藝，故唐人無不工楷法」。宋朱翌(無年代)也說：「唐人無不善書，遠至邊裔書史里儒，莫不書字有法，至今碑刻可見也，往往勝於今之士大夫。」

抄書，「小楷精謹，無一字肆筆」（劉煦等 a，1985，卷 165，頁 4307）、「楷小精真，無行字」（歐陽修等，1985，卷 163，頁 5025）。說他抄書小楷精謹，甚至連一個快筆些的行書都沒有。唐裴行儉工於草書，受命高宗以草書寫《文選》，「帝覽之稱善，賜帛五百段」（劉煦等 b，1985，卷 84，頁 2802）。至歐陽修（無年代）仍讚嘆地說：「唐世執筆之士，工書者十八九，蓋自魏晉以來，風流相承，家傳少習，故易為能也。」

這個以抄書為讀書的儒雅傳統，在面臨雕板印刷術進入經史子集四部圖籍複製技術後，發生了何種衝擊？今天我們理解雕板印刷術，認為它推動了典籍的大量產生，知識的空前普及，藏書文化形成，藏書家輩出，乃至圖書內在結構的變化以及文士創作行為的改變等等，當然，還牽動了整體社會的變革，我們朝向文明曙光騰現來看待此一發展。這樣的觀點是事實。不過，任何舊慣習的被挑戰與新秩序的隱然成風，必定經歷一段迷惘、適應期，不管這新發明在後世看來是多麼毫無疑義。期間有感知、有好奇，有比較、也有質疑，人們會在現實便利的驅使下嘗試、接受新事物，新事物有利有弊，在一開始，沒有人能斷定未來絕對的走向，只能一邊觀望，一邊擇取，靜待潮流的趨勢作出流向選擇。

（二）定本精神的官方雕板印書

唐代的雕印成品，今天知道多屬佛教文獻以及民間曆書、日用通俗類書、字書等內容，印本品質不一，也還未廣泛影響士人階層。五代開始有經書刊印，從宋王溥《五代會要》中那段知名記載：

後唐長興三年二月，中書門下奏，請依石經文字，刻九經印板。勅令國子監集博士儒徒，將西京石經本，各以所業本經句度，抄寫注出，子細看讀。然後顧召能雕字匠人，各部隨秩刻印板，廣頒天下。如諸色人要寫經書，並須依所印敕本，不得更便（使）雜本交錯。（王溥，1988，卷 8，頁 96）

後唐「九經」雕印，實際上是漢代以來官方頒定石經的變化版。它們起因於「經籍失正」（唐·封演，無年代）⁸，影響著儒學，尤其是科舉考試，士子需求正確經書文字，朝廷是以選擇善本，詳加校勘，刻石頒列。歷代石經準漢例，多立於太學堂前，在手寫時代，這已是相當理想的定本辦法。不過，仍受限於距離以及對勘謄正時的未

⁸ 封氏聞見記（唐·封演，無年代）中云：「天寶中予在太學，與博士諸生共論經籍失正，為欲建議請立大唐石經。遷延未發，而蕃寇海內，文儒道消。」

便，五代在六朝、唐以來長期蓄累的碑刻經驗上，加以唐末逐漸風行起來的木板刊印，結合出以木板代石材，雕刻經書文字，並藉由木板刷印功能，成就了雕板印刷的新局。同樣擇選優良經本，同樣發動官方學術校勘，同樣講究書寫、雕工，不同的是，它可以印行天下。⁹

北宋官方大規模校勘刊印經書、史書、醫書、子書、兵書等，太宗淳化五年(994)，詔選官分校《史記》等多部史書，「既畢，遣內侍裴愈賚本就杭州鏤版」(程俱 a，無年代，卷 2)；真宗咸平三年(1000)，詔選官校勘《三國志》等，「五年校畢，送國子監鏤版」(程俱 a，無年代，卷 2 中)；真宗咸平四年(1001)，國子祭酒邢昺等表上重校《周禮》等七經，「凡一百六十五卷，模印頒行」(程俱 a，無年代，卷 2)；真宗景德二年(1005)「詔可。命崇文院檢討杜鎬與奭同校定摹刻」《莊子》(王應麟 a，無年代，卷 55)；景德二年「令雕印放行」《農田敕》五卷(王應麟 b，無年代，卷 178)；仁宗天聖五年(1027)「令國子監摹印放行」校正《黃帝內經》《素問》(王應麟 c，無年代，卷 63)。「和戎偃革，志在古文」的宋真宗，景德二年五月來到國子監「觀羣書漆板及匠者摹刻」，問了國子祭酒邢昺監板數量，獲得的回應是「國初印板止及四千，今僅至十萬，經史義疏悉備」(王應麟 d，無年代，卷 27)。開國不到五十年，僅國子監刻書就達十萬板，成長驚人。

北宋國子監校刻、頒行圖書是官方定本，這些典籍及文獻都與國家文學、武舉考試或攸關民生等十足相關，它的意義等同石經，且是國家學術單位專精學者的心血，不論內容或工藝，都沒有可質疑的地方。宋太宗雍熙三年(986)敕《說文解字》雕造：

說文起於東漢，歷代傳寫，譌謬實多。六書之蹤，無所取法。若不重加刊正，漸恐失其原流。爰命儒學之臣，共詳篆籀之跡。右散騎常侍徐鉉等，深明舊史，多識前言。果能商榷是非，補正闕漏。書成上奏，克副朕心。宜遣雕鐫，用廣流布。自我朝之垂範，俾永世以作程。其書宜付史館，仍令國子監雕為印板。依九經書例，許人納紙墨價錢收贖。兼委徐鉉等點檢、書寫、雕造，無令差錯，致誤後人。(許慎，無年代)

⁹ 五代時期，後唐沒有雕刻石經，稍後長興三年(932)的後蜀廣政元年(938)，蜀主孟昶命宰相毋昭裔在成都督造石經，此石經一直續刻到北宋才完工。北宋、南宋、清代也還刻有石經(毛遠明，2009)，石經在雕板印刷術發達後還進行製作，它的意義是象徵性遠過實際助益。

刊刻《太平聖惠方》序則說：「仍令雕刻印版，遍施華夷，凡爾生靈，宜知朕意」（曾棗莊、劉琳主編，2006，卷 78，頁 407）。而同一時期，官方傳統圖書複製方式——抄寫，進行的規模其實比雕板更大。顯著的事實像真宗咸平二年（999）令昭文館、史館、集賢院等三館「寫四部書二本」（程俱 b，無年代，卷 2 中），當時三館典藏保守估計也有數萬卷。不久後，王宮失火，延及崇文院，即三館藏書，書籍多燬，於是「重寫書籍」（程俱 b，無年代，卷 2 中）。其他詔求士民藏書繕寫以及館閣陸續補寫之書，計太祖、太宗、真宗三朝，有書三萬九千一百四十二卷（脫脫等，1983，頁 5033），寫本是大多數。

北宋初，雕板印刷技術未真正普及，除了少數官方標準本宣示外，宮廷與一般文士仍大抵沿承舊習，長置楷書手抄寫，躬自鈔錄或書工代勞，傳統典範與秩序依舊穩固。然而新技術確實帶來了一些興奮，宋真宗於景德二年（1005）在國子監與邢昺對話，邢昺曾言：「臣少時業儒，觀學徒能具經疏者，百無一二，蓋傳寫不給，今板大備，士庶家皆有之，斯乃儒者逢時之幸也」（李燾 a，1961，卷 60）。真宗十分高興。大中祥符三年（1010），真宗對資政殿大學士向敏中說：「今學者易得書籍」，敏中回以「國初惟張昭家有三史，太祖克定四方，太宗崇尚儒學，繼以陛下稽古好文，今三史、三國志、晉書皆鏤板，士大夫不勞力而家有舊典，此實千齡之盛也」（李燾 b，1961，卷 74）。

另外，也有少數士人開始嘗試這種新技術。宋徐鉉（無年代）《韻譜》後序言：「因取此書，刊於尺牘，使模印流行，比之繕寫，省功百倍矣」。宋張詠（無年代）《許昌詩集》序言：「依舊本例，編為十卷，授鬻書者雕印行用」。宋晏殊《徐公文集》後序：「寶茲遺集，積有歲時，鏤板流行，庶傳悠永」（徐鉉，無年代，卷末）。宋施昌言《唐文粹》後序：「臨安進士孟琪，代襲儒素，家富文史，爰事摹印，以廣流布。觀其校之是，寫之工，鏤之善，勤亦至矣！……今是書也……無煩簡札而坐獲至寶……」（姚鉉，無年代，卷末）。以上引文文字中多見特別與繕寫作比較，說明這是一種新型態的方法。

（三）雕板印書的大眾化與庸俗化

隨著宋代社會的改變，教育發展與科舉制度等直接需求促使下，雕板印刷複製圖書便利，自此很快發展起來。在過渡階段，宋初編集《文苑英華》，曾批評所收唐盧

肇作品，「盧肇有集印行，大抵不若寫本之善」(李昉，無年代，卷 10)。宋李燾《續資治通鑒長編》據王暉《百一編》，也記載了這個時期的一段故事。

國家三年一修郊禮，必有肆赦。寇萊公嘗議模印以頒四方，為眾所沮乃止。其後外郡覆奏，赦書字誤。沂公始舉寇相之議，令刑部鎖宿雕字人模印頒行。因之，日官乞每年頒歷日，亦雕板印行。舊每歲募書寫人，所費三百千，每模印則三十千。仍有沮議曰：「一本誤則千百本誤」。沂公語之曰：「不令一字有誤，可矣」，自爾遵行不改(李燾，1961c，卷 61，葉十)。

寇準在真宗朝為相時，對於頗為頻繁頒布的赦文有過建議，不如雕板印行，廣發地方。但赦文是何等重要的朝廷文書，一向以來由「院吏筆札精熟者書寫」(李燾 d，1961，卷 61，葉九)，且經官員校讀，不容差錯。雕印的想法沒有採行。後來仁宗朝王曾當國，地方反應赦書竟有誤字。王曾重提當年寇萊公建議，不過短短一、二十年，情況變化了，不但即刻施行，連職掌曆書的單位也馬上跟進。一個難免的字誤被放大，成了仁宗朝政府官員終能突破傳統、擁抱新技術的理由。加上所費減省十倍的記錄，不難看出他們的一絲欣喜，省錢又方便。即使仍有異議，王曾也輕描淡寫帶過。可見當時技術已然純熟，士庶使用印本已具信心。而「一本誤則千百本誤」，則也確實踩住雕板印刷的最大痛處。

宋蘇軾(無年代)曾追憶自己見過一老儒，說其人「少時欲求《史記》、《漢書》而不可得，幸而得之，皆手自書，日夜誦讀，惟恐不及」。也不過二、三十年時間，蘇軾的時代已是「近歲市人，轉相摹刻諸子百家之書，日傳萬紙」(蘇軾 b，無年代，卷 32)，雕板印書已是坊間市人習為之事。然而，「學者之於書，多且易致，如此其文詞、學術，當倍蓰於昔人。而後生科舉之士，皆束書不觀，遊談無根，此又何也？」(蘇軾 b，無年代，卷 32)，如此迅速，大量刻本圖書改變了士風。蘇軾也屬過渡階段的士人，但對後輩則不無感慨及疑惑，為什麼書多了反而不再讀了？

雕板印書風起水湧，鬻書者多唯利是圖，有銷路則不顧其他。宋李覲在仁宗皇祐四年(1052)的慘痛經驗：

覲慶曆癸未秋，錄所著文，曰《退居類藁》十二卷。後三年，復出百餘首，不知阿誰盜去。刻印既甚差謬，且題《外集》，尤不韙心，常惡之，而未能正。(李覲，2004，卷 25，頁 176)

盜刻、謬題、粗製濫造、錯誤內容，作者為之氣結，卻也毫無辦法，印本早流行天下。宋蔡襄(無年代)在仁宗時寫過《茶錄》二篇，「後知福州，為掌書記竊去，

藏稿不復能記。知懷安縣樊紀購得之，遂以刊勒，行於好事者，然多舛謬」。身為作者的蔡襄無法忍受，「輒加正定，書之於石，以永其傳」（宋·蔡襄，無年代，卷 25），乾脆選擇傳統刊布方式一刻石，來正定作品內容。直到南宋寧宗嘉泰四年（1204），陸游還是有同樣的感嘆。他曾獲一本歐陽修名著《集古錄》，「刻畫精緻，如見真筆。會有使入蜀，以寄張季長。及再得之，纔相距數年，訛闕已多。知古人欲傳遠者，必託之金石，有以也」（陸游 a，無年代，卷 30）。圖書文籍是容易取得了，但輾轉流傳翻刻，過程不知有多少輕率改易，根本浮濫難以控制，錯謬也隨之張揚擴散。這使得部分退轉的聲音產生了。

宋孔延之早在神宗熙寧五年（1072）《會稽掇英總集》原序中稱，古代《詩》、《書》多闕亡，是因「編脫簡落」，後世文章「藏之名山，副在緗帙」，有珍重的典藏與保障性地鈔寫存副。可是到了鏤板與石刻，板片一旦不存，岸谷一旦變易，就什麼也不見。何況戰火、人禍頻仍，木板與石頭都難長存。是以：

題之板不如刊之石，刊之石不如墨諸紙。苟欲誦前人之清芬，搜斯文之放逸，而傳之久遠者，則紙本尚矣。（孔延之，無年代，卷 1）

孔延之的說法不盡合理、周全，但從文中舉唐白居易、元稹兩人詩作比較看來，他的意思是像白居易那樣能自己寫存作品並多次編集，詩歌自可完整保留，元稹悠然隨意，詩作再好也不知所終。間接強調了寫本的經典性。此外，宋司馬光談史書「今國家雖校定、摹印正史，天下人家共能有幾本？久遠必不傳於世，又校得絕不精，只如沈約叙傳，差却數板，亦不寤，其它可知也」（司馬光，無年代，卷 62）。宋朱彧（無年代）《萍洲可談》記載了那個州學教授看了福建本出錯考題的知名笑話。¹⁰ 錯字、闕板、跳頁，這些破壞文本內容的狀況，隨著雕印手工業的昌盛，愈來愈嚴重。而宋葉夢得的一番話，尤令人印象深刻。

唐以前凡書籍皆寫本，未有模印之法。人以藏書為貴，不多有而藏者精於讎對，故往往皆有善本。學者以傳錄之艱，故其誦讀亦精詳。五代時馮道奏請，始官鏤六經板印行。國朝淳化中，復以《史記》、前後漢付有司摹印。自是書籍刊鏤者益多，士大夫不復以藏書為意。學者易於得書，其誦讀亦

¹⁰ 姚祐，元符初為杭州教授。堂試諸生，出易題。「乾為金，坤亦為金也」。蓋福建本書籍刊板舛錯，坤為『釜』，脫二點，故姚誤讀作『金』。諸生疑之，因上請。姚復為臆說，而諸生或以誠告。姚取官本視，果『釜』也，大慙。曰『祐買着福建本』。升堂自罰一直，其不護短如此。

因滅裂。然板本初不是正，不無訛誤。世既一以板本為正，而藏本日亡。

其訛謬者，遂不可正，甚可惜也。(葉夢得，無年代，卷8)

葉夢得幾乎總結了雕印以來出現的所有問題，讀書人不再藏書、讀書與精求內容，完全以市面上輕易獲得的普及本為準，而商業化的版本不一定是最好的本子，但因流通廣，使用率高，寫本也就從無人知曉到不知所終。麻煩的是，印本的錯誤就再也難以更正而不斷傳布下去了。葉氏含蓄地說這很可惜，但事實上應說這是圖書典籍的另種「浩劫」吧！

到了南宋，雕板印刷已普及並風行天下，刊印帶來的裨益自不必說，只要看看南宋雕板印書數量的爆增就可明瞭。中央國子監刻書、地方官府刻書、士人刻書以及形成品牌的書坊刻書，琳瑯滿目。《宋史·藝文志》記載有宋一代官方圖書數量，北宋一百六十七年，在前朝基礎以及歷任君王建立下，最後總數有七萬三千多卷。而南宋經歷靖康之難，典籍蕩然，到了臨安從頭開始，在寧宗之前的短短七十年，已擁書四萬四千餘卷（脫脫等，無年代，頁5033）。這還不計民間圖書，著名的南宋末年藏書家陳振孫，個人所藏即達數萬卷，所著書志大量記錄刻本書（張茂鵬點校，1983）。¹¹ 刻本湧現，造成的問題當然也愈多，對它的糾舉是綿延不歇。

宋周必大（2004）《文苑英華》序說：「國初文集雖寫本，然讎校頗精，後來淺學改易，浸失本指。今乃盡以印本易舊書，是非相亂」。這裡的「淺學」甚至不僅坊間手民，而是連宮廷校書員都散漫應付，避諱不辨，隨意改書。宋樓鑰曾獲一部傳說華陀所著《中藏經》，閩中倉司刊本，由於未讀過此書，覺內容難以盡信，「差舛難據」，於是求教當世名醫陸從老。陸醫驚訝於市面上早有版行，取家中秘藏寫本出示。樓鑰借回校對：

乃知閩中之本未善，至一版或改定數十、百字，前有目錄，後有後序，藥方增三之二。閩本亦間有佳處，可以證陸本之失，其不同而不可輕改者，兩存焉，始得為善本。(樓鑰 a, 1979, 卷71)

兩個本子內容頗見差異，人命關天的藥方，刻本多了三分之二。在沒有更多本子可以讎校下，樓鑰只能保留地兩存焉。又在〈跋《春秋繁露》〉文中，樓鑰有另種經驗：

¹¹ 近年惟直齋陳氏書最多，蓋嘗仕於莆，傳錄夾漈鄭氏、方氏、林氏、吳氏舊書，至五萬一千一百八十餘卷（張茂鵬點校，1983）。

《繁露》一書，凡得四本，皆有高祖正議先生序文。始得寫本于里中，亟傳而讀之。舛誤至多，恨無他本可校。已而得京師印本，以為必佳，而相去殊不遠。（樓鑰 b，1979，卷 77）

看來「一本誤則千百本誤」的憂慮，絕不是一個誇張假設。宋陸游在寧宗慶元六年（1200）跋《唐盧肇集》：

子發嘗謫春州，而集中誤作「青州」，蓋字之誤也。題〈清遠峽觀音院詩〉，作「青州遠峽」，則又因州名而妄竄定也。前輩謂印本之害，一誤之後，遂無別本可證，真知言哉。〈病馬詩〉云「塵土卧多毛已暗，風霜受盡眼猶明」，足為當時佳句。此本乃以「已」為「色」；「猶」為「光」，壞盡一篇語意，未必非妄校者之罪也，可勝歎哉？（陸游 b，無年代，卷 28）

陸游也無法確定是誰竄改了作品中的文字，但這多半不是文士們所為，而是所謂書坊編刻。坊肆編輯者雖非白丁，文化素養卻不見得具水準，在缺乏學識基礎加上自以為是的情況下，一誤牽連再誤，原作遂面目全非。類似的例子在宋代不勝枚舉。當然，文本有誤，不是印本的專利，寫本也有同樣困擾，只是寫本影響層面與程度相對有限。陸游便曾灰心地說：

近世士大夫所至，喜刻書版，而略不校讎，錯本書散滿天下，更誤學者，不如不刻之愈也，可以一歎！（陸游 c，無年代，卷 26）

連士大夫們所刻的書都不免貽誤。又傳為蘇軾所著的《東坡志林》，也記載了一段充滿憤慨的文字：

近世人輕以意改書，鄙淺之人，好惡多同。故從而和之者衆，遂使古書日就訛舛，深可忿疾。（蘇東坡，無年代，卷 5）

寫本時代書不易得，唐杜暹家藏萬卷，「清俸買來手自校」（周輝，1981，卷 4，頁 21），結合宋人普遍概念，事實應該相差不遠。他們具學識，能讀書，勤讎校，學問與擁書、藏書是同一件事。宋代基本上社會型態就不一樣了，科舉考試向更多數平民百姓開放，基礎教育較前發達，雕板印書事實上也是應運而生。新技術擺脫純人工，神奇地使書籍轉瞬化生千萬。它的從業者不再是深習墳典的文士，而是寫字、刻字與刷印工匠，加上或多或少具知識的坊肆主持者。書籍商品化、市場化的程度更深了，擁書與讀書也正式分化，一種「庸俗化」的氛圍與感受，在過渡與接受的過程中瀰漫。

然而，雕板印書真的不如寫本？宋劉跂為趙明誠作《金石錄》序，早有一番持平之見：

昔文籍既繁，竹素、紙札轉相謄寫，彌久不能無誤。近世用墨版摹印，便于流布，而一有所失，更無別本是正。然則謄寫、摹印，其為利害之數，畧等。(劉跂，無年代，卷6)

寫本的舛誤不見得少，否則唐代士人也不必強調廣求善本以校讎。印本的問題在交由營利者來再製圖書，業者可能不懂文意卻擅改古書，可能不辨利害而濫刊文籍，然整體製程比過去短，傳播卻是過去完全無法匹敵的，為利世人之處不待多言，為害學術或其他之處則教人不安。而這些幾無控制辦法。宋代有多次官方禁令刊印圖書，如哲宗元祐五年（1090）禮部建請「凡議時政得失、邊事、軍機文字，不得寫錄傳布。本朝會要、國史、實錄，不得雕印。違者，徒二年」（李燾 e，無年代，卷 445，葉十一）。就是「其他書籍欲雕印者，納所屬申轉運使、開封府。牒國子監選官詳定，有益於學者，方許鏤板」（李燾 e，無年代，卷 445，葉十一）。這個建議是來自翰林學士蘇轍說過「奉使北界，見本朝民間印行文字，多已流傳在彼」（李燾 e，無年代，卷 445，葉十一）。

奴役兩年，罪責非輕。其他禁曆書、印賣禁書、盜版、科考小抄等等禁書令（周寶榮，2003，頁 133-176），都反映出雕板印刷已然涵蓋所有圖書文獻領域，且滲透到全天下每個區域，人們需求它帶來知識、普及文化、便利生活，即使印刷品內容可能有誤、或許不當，但也再不能回到侷限上層知識分子的時代。庸俗化了，又如何？而時至明代，雕板印刷更加發達，庸俗化程度再飆升，明、清對宋代刻本珍若拱璧，一葉百金，這些可不乏當年曾被訾議的版本。而那些精緻罕傳的寫本呢？沒有搭上雕印列車，隨著秘藏凋零、孤本沉埋，永遠消失於世間了。

三、當今古籍數位化所面對的成效質疑

上述雕板印書發展與士人接受情況描述，如果把雕板改成今天的數位化，雖然分明是完全不同的時空與背景，但發生的狀況與人們的感受，竟有奇異地類同，許多層面上頗令人會心。當然，雕板印書的挑戰小多了。首先，它們的載體沒有改變，是複製技術改良了。而這技術的全部要件，在過去早已具備並成熟。刻書前以人工書寫，上板後刻工雕字，刷印、裝訂，完成圖書。人工書寫部分幾可完全沿承唐代的書寫習慣，事實上在雕板印書的全時期發展中，刻本文字一如手寫風格，一板中有同樣的字，

往往使用異體字，不令篇幅中重複過多相同字形，就是常見字也喜用異體，這在今天只有書法藝術還保留此習慣。異體字在刻書中不成問題，古字、避諱，都做得到。而雕刻文字，六朝、隋唐以來發達的碑刻技術，硬質的石材與相對軟質的木板比起來，難易立見，工藝技術也沒有障礙。其次，在分工上，只有選書及單純的寫、刻、刷印等純技術人員，然後就是使用的讀者，中間沒有更多技術上的轉接。因此，宋代也不乏學者文士投入刻書，在刻書前善選優良本子，過程中要求質量，以及像國子監及地方官府等，具團隊能力、條件來堅持刻書品質的，都可以使雕印達成令人滿意的成果。

「缺字」是古籍資訊處理從頭到現在都存在的困擾。中文字本身有複雜性，量大且面貌多樣，如歷代數變的字體、異體字、通假字、罕用字、避諱字等，就是普通用字，數量也超過電腦處理能力。從過去以來，各古籍資料庫為求突破，或自行開發系統、進行造字、置放圖檔或權宜拼字等，今天的問題已沒有製作之初嚴重，條件較好的團隊解決能力也較佳。¹² 然缺字又引發另一個課題，數位文本要不要完全再現紙本面貌，一字一形不苟？！異體字有沒有堅持保真的必要？因為這不單純是缺字問題，還牽涉檢索，其實就算電腦打得出來的字，也很可能因輸入時字形選擇不同、內碼有異而影響之後檢索，那就從根本上動搖了數位化的效益。

古籍數位化面臨全盤再現古書原始面貌的任務，可是諸多難題不易解決。雕板不脫人為書寫，但數位化卻是媒材的徹底改變，受限於工具，無法以直觀的方式處理。最直觀的方式便是掃描成圖檔，但那又完全喪失資訊化意義——強大便捷的檢索與知識挖掘。若加上不論任何載體圖書都必定發生的錯誤以及版本問題，數位化既費心力又耗金錢，最重要的，問題解決有限。問題難解，因為電腦資訊是一獨立學科，不僅是工具，非為服務古籍而存在。古籍要利用它，必須適應它的原理與要求，才可能驅動它協助特定需要，這已遠非文史研究人員所能掌握，是以分工上多了一個資訊工程師的角色。資訊工程師少有能懂古代文史專業者，他們也只能依技術潮流提供工具建議或儘量達成文史專家的要求，而能深入的不多見，因為文史專家也提不出所以然的意見，這一切尚不計文史專家有沒有資金與人力奧援投入。

¹² 如北京「國學網·國學寶典」的製作，主持者尹小林先生即宣稱：「Unicode 有 7 萬多字，再加上一些造字，漢語存世古籍的字形檔問題基本上解決了」。筆者以該資料庫提供的免費首頁查詢進行檢查，「國學寶典」中仍出現有拼字情況，如「爛喉<疒丹>痧輯要」；又，資料庫中異體字多改為今通用字，或許這便是字形問題縮小的原因。

古籍數位化的問題相當繁多，本文僅從幾個方面來進行討論，主要是以一個文史研究者使用心得的立場，以及初步參與製作的經驗，希望提出在過程中關心的問題與面臨的困擾，以體現當今古籍數位化建置成效致疑的情況與原因。

(一) 大型古籍全文資料庫的驚豔與驚詫

古籍數位文本今天還不是拿來閱讀的，主要功能仍在資料查找。古籍利用，較大的客群仍然是學術研究，普羅大眾需求有限。因此，以學術要求來說，它被期待擁有紙本的原始內容、類同形式，具備圖書館式的豐富典藏，清晰完善的書目著錄，優良的版本，正確無誤的文字，然後又可以在網上隨心使用，能省去跑圖書館的辛勞，毋須架上找書的麻煩，輕鬆以輸入檢索掌握所需要的圖書與內容，並且當引用時，無異眼見的紙本作品。

我們從幾個收錄豐富的大型古籍資料庫來觀察，北京愛如生「中國基本古籍庫」、北京國學網「國學寶典」、大學數字圖書館國際合作計畫「古籍」部分、中央研究院「漢籍電子文獻」。「中國基本古籍庫」收錄超過一萬種古籍，數量豐富，種類均勻，內容核心，有全文檢索，也附一至二種原書圖版，是相當吸引人的工具資料庫。不過，它的問題有三，一為全文文字錯誤過多，已遠遠超過可容許範圍。有直接缺了幾葉古書版面，也有數不清的錯字。使用者目前可從資料庫所附工具輕鬆下載片段全文文字，但幾乎不敢逕行引用。二為圖版無法與全文連動，不能相信全文，至少可比對圖版，但圖版瀏覽閱讀不便，耗去的時間不見得少於紙本翻閱；且圖版時見裁切過甚，版心葉數多半難以察明。三為全文去除序跋、目錄、附刻及識語等相關內容，是嚴重的闕損。其他有版本選擇未佳及書目著錄不盡清晰等，相對來說還是末節。

國學網「國學寶典」依官網所見，2013 版收錄 4,903 種古籍，且每月都持續新增（北京國學時代文化傳播有限公司，無年代）。以網站所提供之資料及免費試用後之了解，重要典籍涵蓋完善，主要是全文檢索資料庫，最大特色為具標點，且有繁簡字切換，不過有繁簡字直接系統轉換的別字問題及未見提供布林邏輯式高級檢索。若試以「雜」與異體「禿」兩字輸入查詢，試用所見是不相同的典籍內容。程毅中(2013)曾為文指出該社出品《中國歷代筆記》U 盤，存在收書版本說明不清、刪去序跋、內容選擇不精、作者名未核實、標點錯誤以及未良好利用後人成果等，認為數位化古籍整理的學術素養應該加強。

大學數字圖書館國際合作計畫「古籍」部分，是一深益國學研究者的工具。它只要註冊，「古籍」部分不設使用權限。依網站所見，古籍數量近 27 萬冊（北京國學時代文化傳播有限公司，無年代），其中有重複內容或不同版本的相同作品。此資料庫為掃描圖檔，可見圖書原始面目，且畫面十分清晰，但當然無法檢索內文。能夠節省往返圖書館時間，已是對學者們的莫大造福。可惜它的書目資料建置不全，因所收錄的作品大抵是新式影印本，理應有版權訊息與頁碼，而資料庫中呈現未完整明確，對引用來說障礙鉅大。是以網站首頁也祭出獎賞，鼓勵讀者協助「定位描述不佳的圖書」。

臺灣最大的古籍文本製作—中央研究院「漢籍電子文獻」，據網站首頁顯示已收 800 餘種古籍（中央研究院歷史語言研究所，無年代）。此資料庫為半開放式，一般使用者僅能看到少部分內容。資料庫可檢索、可瀏覽作品全文，部分有圖檔相應。進階檢索提供布林邏輯選擇，異體字與同義詞也可以勾選啟用。它的優點在內容正確，缺字處理十分盡力，相關書目訊息也清楚，具引用條件，不過未全面標點。而它最大的缺點即收錄太少，尤其開放的數量微薄，對學術研究協助有限。

綜合以上情況，今天的古籍資料庫堪稱已完成核心古籍的處理，且具備多方面功能與細膩加值，提供學術研究查詢與利用，是教人驚豔的新時代利器。它們確實刷新了學術研究的方法，解除了學者耗費在圖書館翻查的大量時間與精神，甚至大幅超越過去最為用功的成果。但它還是存在頗嚴重的內容正確與否問題，以及檢查查全、查準的問題，使它們迄今尚未能令學術界全面接受，至少是要求嚴謹的專家還存在疑慮。不過，當資料量遠遠大過畢生辛勤所能掌握範圍，即使有錯誤、不完整，事實上也只好接受。可惜像「中國基本古籍庫」與「國學寶典」這些大型且加值豐富的資料庫，使用門檻很高，費用不便宜，學界中能利用者有限；「漢籍電子文獻」內容不多卻也保護過度，稱不上實質開放。這大大降低了它們的影響性，也使古籍數位化成就再怎麼喧騰，仍是一座孤島。

（二）取代落空的數位化古籍叢書目錄

缺字干擾了古籍製作與檢索，但挑戰遠不僅此。筆者數年來進行古籍叢書目錄的數位化與資料化工作，經歷複雜的各式問題。由於紙本叢書目錄查檢不便，也為了較全面地整理古籍叢書，遂進行幾部主要目錄的整合並以資料庫型態呈現，希望能全面取代紙本。叢書子目繁多，不少是生僻的作品，異體字與罕用字常見。為再現紙本全

然面貌，起初即完全按所見打字與校對。缺字部分由於有系統工程師專利開發中文字庫，於 Unicode 以外再直接補入九成，最後一成方採造字。然字形問題未解決，因輸入與校對都為眾人合力完成，過程中大家對紙本所見字的認知與選擇偶有差異。如「冊」與「册」，在用字看來，兩者無異，但之後的檢索就出現問題，電腦以二字內碼不同，判為不同的字，檢索結果彼此睽違。以一般較常選用的「冊」輸入，大範圍漏失以「册」建置的內容，類似情況很多，無法逐字校正。

其次，為開發內容資料，在工程師協助下，利用文字編輯器進行叢書名、編著者、版本、典藏地、子目書名、子目作者等 XML 語法標示 (Markup)，以方便未來資料連結與程式運算。如：

```
<pb n="91"/>
叢書分類詳目
<cat level="1">彙編叢書</cat>
<cat level="2">雜纂類</cat>
<_ id="1001.百川學海二十卷"/>
<coll j="20">百川學海二十卷</coll>
  宋<pr>左圭</pr>編
    <ed>明嘉靖十五年 (1536) 鄭氏宗文堂刻本</ed>
  總卷一
    <ti j="1">聖門事業圖</ti>一 JUAN 宋<pr>李元綱</pr>撰
    <ti j="4">學齋佔畢</ti>四 JUAN 宋<pr>史繩祖</pr>撰
    <ti j="3">釋常談</ti>三 JUAN
  總卷二
    <ti j="3">中華古今注</ti>三 JUAN 五代<pr>馬縝</pr>撰
    <ti j="1">漁樵對問</ti>一 JUAN 宋<pr>邵雍</pr>撰
```

但，紙本目錄著錄時體例不完全一致，又或者為簡省，但到了講求精確的資料庫中，參差情況未免。如

```
<coll>清獻堂全編</coll>
  (清) <pr>趙佑</pr>撰
    <ed>清乾隆五十二年 (1787) 刊本</ed>
    <ti j="8">詩文集</ti>八卷
    <ti j="2">尚書質疑</ti>二卷
```

<ti j="6">尚書異讀考</ti>六卷
 <ti j="2">草木疏校正</ti>二卷
 <ti j="10">春秋三傳雜案</ti>十卷
 <ti j="4">讀春秋存稿</ti>四卷
 <ti j="12">詩細</ti>十二卷
 <ti j="11">四書溫故錄</ti>十一卷

使用者若想找「清獻堂詩文集」，輸入此六字查不到資料，然而使用者未必知道只能用「詩文集」三字來查。又如：

<ti j="16">滿州金石志</ti>六 JUAN<ti j="2">別錄</ti>二 JUAN<ti j="1">補遺</ti>一 JUAN<ti j="1">外編</ti>一 JUAN<ti j="1">校記</ti>一 JUAN

「別錄」、「補遺」、「外編」、「校記」只有緊隨「滿州金石志」書名才有意義，獨立標示與否出現為難。以上從字形到內容的問題，在紙本查檢中全然無誤無礙，資料庫費了那麼多功夫，卻掛一漏萬，令人氣餒。而且資料庫是虛擬符號，使用者幾乎難以發現資料的遮蓋。另外，紙本目錄已做到子目類編，這部分資料庫還完全趕不上，等於二分之一的功能缺損，離取代紙本尚遙遠。

（三）統計量化在文史研究上的意義

數位化的強項之一即運算，儘管一向以來文史討論以質性研究為主，強調現象觀察背後的意義分析與詮釋，不以數量為指標，但在工具潮流中也不免嘗試新技術、新方法。原本感發性的詩歌風格分析，採用詞頻運算來觀察用字與詞語習性，得出量化的認知。但問題來了，對於隱含義如何檢索？於是有所謂「語意概念索引」的設計。如臺灣「網路展書讀」網站（<http://cls.hs.yzu.edu.tw/home.htm>），主持人羅鳳珠進行詩歌內容的文字切分與性質標示，提供讀者選擇。

唐代詩人行吟地圖：
李白、杜甫、韓愈

回首頁 ▶ 全文檢索 ▶ 全文索引 ▶ 唐代地名資料庫 ▶ 唐代詩人行吟地圖唐代地名地理資訊系統 ▶ 系統簡介 ▶

詩文檢索

作者姓名： 請選擇 李白 杜甫 韓愈

作品文體：

作品繫年： 選擇全部 西元 帝號年號 作者年紀 「選擇全部」則不用輸入，若選擇其他選項，請記得輸入條件。

總類：

語義概念索引：
大類：
小類：
元素：

可搭配詩文欄位使用，更詳細檢索出您欲了解的詩文辭彙。

詩題：

詩序：

詩文：

詩文寫作地：
路：
州：
縣：
地名：

詩文篇名地：
路：
州：

圖 1 「唐代詩人行吟地圖：李白、杜甫、韓愈」詩文檢索首頁。圖片取自 http://cls.hs.yzu.edu.tw/TWSLDH/TWSLDH_Srchmain.aspx。國科會數位典藏國家型科技計畫版權所有(NSC100-2631-H-155-002)。

然而，語義分析無法大量、大規模進行，分析方式與類目用語也是一大難題，規劃者與讀者的認知不容易有絕對共識。那麼，統計量化還是有不完全的問題。就是不牽涉特定用語的全文語法分析，還是可能出現爭議。知名的《紅樓夢》後四十回作者之爭，先後有以電腦運算一百二十回全文語法，結果出現後四十回為曹雪芹原著及他人所續兩種答案，問題再度回到原點（陸宇傑、許鑫、郭金龍，2012）。

此外，類似性質問題，如臺灣學者黃一農提出過「e 考據」名詞，即大量從資料庫中蒐羅某一特定詞彙，進行相關史學考證。資料與證據的數量史上最高，但也因龐雜，還是免不了錯認誤解（吳夏平，2012）。那麼，量化在文史研究上真的是一個新天地？它為我們帶來何種新觀點、新知識？傳統以來的因果關係解釋與背景前提分析，是不是值得為量化掩蓋？

古籍數位化的發展，人人叫好，學界樂於有便捷工具可使用，但看看今天多數學者作品中，廣泛表示引用資料庫的仍少見，嘗試量化研究的更稀少，對資料庫的疑慮不言可喻。

四、古籍文獻的「大數據」遠景

臺灣資訊化先驅謝清俊（2014年）在中國科學院第10屆「兩岸三院資訊技術與應用交流研討會」，發表「淺談人文資訊學的回顧與省思」簡報。他在人文資訊學發展緩慢議題中提出省思：「目前，人文學者利用數位工具協助研究，還泰半停留在數據或資料處理的層面。這還是1980年之前的一般水平啊！遲滯了三十年吶！」謝教授談的是資訊技術能否及如何協助人文學者做「理解」與「意義」的研究，他深深了解學科性質的不同、追求的不同，而期許資訊領域不要畏懼困難，努力嘗試達到對人文學科的影響。因為，資訊科技「對每個領域都提出了對問題的新看法，並提供了對問題詮釋的新角度和解決問題的新方法，進而建立了對問題理解的新模式和新理論。」人文學科也不可能自外於這場改變。謝教授沒有責備人文領域的意思，是作資訊端的反省，但文史研究者難道不需要思考，為什麼我們落後了三十年？新看法、新角度、新方法、新模式與新理論，連五個新，文史界要持續作壁上觀嗎？

2012年牛津大學網路研究所教授麥爾荀伯格 Viktor Mayer-Schonberger（）以及《經濟學人》雜誌資料編輯庫基耶（Kenneth Cukier）合寫了一部《大數據》，宣告「『數位革命』之後，『資料革命』登場：巨量資料掀起生活、工作和思考方式的全面革新」。此書一出，很快成為網路書店第一名暢銷書，被喻為十年一見的改變觀念之書（麥爾荀伯格、庫基耶，2013／林俊宏譯，2013）。數位文明已然再度進化，從原先的大規模資訊轉換，到現在生產出人類文明的新資料品種，而且數量龐大到自己可以建立觀點、發言說話。文史界的資訊課題，似乎不容再忽視了。「大數據」的全部概念或許不盡適用於文史數位化討論，但它提出的思考與方法理念，值得我們借鑑。

（一）技術突破與開放文本

大數據的基本精神在發現與創意開發鉅量資料的價值，提醒我們一個資料導向的社會與世界即將到來。即「當一切成為資料」、「量化這個世界」，所掌握的訊息可展延為「無窮無盡」的用途（麥爾荀伯格、庫基耶，2013／林俊宏譯，2013，頁113、133）。¹³ 資料化與數位化還不同，數位化僅是將文獻錄為數據或掃描，它們必須被

¹³ 「大數據」書中所談其實多為被動性蒐集資料，與我們文史數位化尚須主動建設基礎，階段不同。但它闡釋的唯有全面資料化以後，才能進行各種創意觀察，發現隱藏的認識，提供人們解決問題，適用文史研究參考。

進一步詮釋及判斷，才能成為可分析的資料，也才可以成為資訊人眼中真正利用來「做研究」的工具。

「資料化」不是新名詞，就如同 1950 年代史學家嚴耕望（1956）為研究唐代士風與黨爭，細密考證了唐代二千六百八十餘任共 1,116 位尚書省左右僕射、左右丞、六部尚書與侍郎官員的籍貫、出身、任職等資料，編為圖表；為寫作《唐代交通圖考》，從《全唐詩》找出 1,000 條以上的相關佐證（嚴耕望，2008）。這裡都看到量化的研究法。古籍數位化資料也要朝這樣的方向加值，進行數位文本的標示，讓一次整理成果，提供無數嚴耕望的創意組合取用，而或許省下嚴耕望那種畢生孜孜矻矻、毫不懈怠喘息的資料爬梳工作。文史研究者資料爬梳工作不僅在獲取材料，也從中奠定對專業的認識與見識，不是區區資料蒐集可以論價值的。但，如果每一個有志研究者都要重新去梳理材料、提取關鍵詞彙，重複耗費時間精力，在數位時代是一種消耗與浪費。更重要的，如果對象只是一部總集、幾部史書與政書，那還在控制範圍，而若跨越的時間與空間超過人的可能呢？

資料庫進行全文檢索，只能找到完全符合檢索詞內碼的詞彙或單字，無法智慧判斷可能的相同資料，其結果是粗略的。但進行像 XML 語法標示，把材料中的詞彙與性質定位且定義出來，則不僅可令檢索時完整無誤，數位文件還可進行大串連，真正地發揮數位化功能，跳脫電子版本的掙板角色（謝清俊，2007，2008）。目前不少個別製作採 excel 檔欄位建置，交由資訊公司轉入資料庫，這樣的數位化未來若無人持續維護，資料庫是停滯的，資料是封閉的，不便再加利用，也不能匯流成資料大軍。

XML 語法標示，如筆者建置之「古籍叢書資料庫」：

```
<_ id="1279">枕經閣叢刊（一名太平陳氏枕經閣叢書）"/>
<coll>枕經閣叢刊</coll>（一名<coll type="alias">太平陳氏枕經閣叢書</coll>）
  <pr>陳樹鈞</pr>編
    <ed>清宣統元年（1909）太平陳氏自刻本</ed>
  <ti j="1">菌譜</ti>— JUAN    宋<pr>陳仁玉</pr>撰
  <ti j="1">讀書錄存疑</ti>— JUAN    宋<pr>潘音</pr>撰
  <ti j="1">田間書</ti>— JUAN    元<pr>林昉</pr>撰
```

給予每一筆叢書如同身分證的編號，說明別名、編者、版本以及子目書名作者資料，有了這些標示，未來可隨心運算其間訊息。此項工作需要文史專家投入，只有專

業才能明瞭這些材料的意義，得以正確標示並規畫未來功能。尤其重要的是，制定一種古籍標示的凡例，如何斷詞、如何突顯各種詞彙。如方志的標示：

<ti j="10">雍錄</ti>十 JUAN 宋<pr>程大昌</pr>撰
 <ti j="4">元河南志</ti>四 JUAN 清<pr>徐松</pr>輯
 [至元] <ti j="6">齊乘</ti>六 JUAN<ti j="1">附釋音</ti>一 JUAN 元<pr>于欽</pr>纂
 釋音<pr>于潛</pr>撰
 [淳熙] <ti j="10">新安志</ti>十 JUAN<ti j="1">附錄</ti>一 JUAN 宋<pr>羅願</pr>纂

考量讀者檢索時會採用的詞彙，進行標示選擇。但仍有可能如「元河南志」，讀者未必輸入完整，在許多須首字符符合的檢索系統中，此資料很可能漏失。已開發之模糊檢索功能，即使輸入的詞彙是不完整或跳離在上下幾行的，系統都可以計算出符合比例，提供選擇呈現。古籍資料要全面進行標示，乍聽下不容易，但事實上今天已有許多軟體工具可輔助，文史專家只要管控那些電腦所無法判斷與理解的部分。如法鼓文理學院所建置的「佛學規範資料庫」建立之「知識規範檔」(authority)。如圖 2 此人名部分，為每一位重要佛學人物建立身分規範碼，包含各種相關知識並多語種內容。完善此規範檔，未來可以利用在數位文本的自動標示，不論文本中出現的是本名、字號或各種習見稱號，都可以被規範、標示出來，也就可以完整檢得。

人名規範檢索

Buddhist Studies Person Authority Databases (Beta Version)

輸入人名(Name) Query

部份符合 完全符合

輸入權威碼(Authority id) Query

輸入舊編碼(Old id) Query

輸入註解(Note) Query

生卒年搜尋

您還沒選擇時間

鳩摩羅什(jiū mó luó shí)

別名 Alternative Names :
 [正體中文] 什(shì), 童壽(tóng shòu), 什師(shì shī), 什公(shì gōng), 羅什(luó shí), 鳩摩羅耆婆(jiū mó luó qí pó), 兜摩羅什(jiū mó luó shí), 鳩摩羅什婆(jiū mó luó shí pō), 拘摩羅耆婆(jiū mó luó qí pō), 兜摩羅耆婆(jiū mó luó qí pō), 鳩摩羅什(jiū mó luó shí), 童受(tóng shòu)
 [梵文 (羅馬轉寫)] Kumārajīva

規範碼 ID No. : A001583
 舊編碼 : B1757

分類 Category : 譯經
 生年 Birth : +0344-01-01 ~ +0344-12-31
 卒年 Death : +0413-05-29
 朝代 Dynasty : 後秦
 生卒年查詢字串 Birth-Death date Quote : 卒於：後秦，高祖，弘始，十五年，四月，十三日，70歲
 生年出處 Birthdate Reference : 依記錄推算。
 卒年出處 Deathdate Reference : 癸丑之年，年七十，四月十三日薨於大寺。(T52n2103_p0264c18)
 編號 From : 龜茲(PL000000047823)
 塔 / 靈寂地 : 罽賓寺(PL000000042857)
 塔 / 靈寂地 註解 : 鳩摩羅什舍利塔位於陝西省西安市西南三十餘公里戶縣的草堂寺內。(http://baike.baidu.com/view/1271250.htm · 2010.12.09)
 性別 Gender : 男
 是否為歷史人物 Historical Person : 是
 註解(簡易) Notes(simple) : 中國佛教史上四大譯師之一，三論宗之祖，七歲從母入道，初習《阿毘達磨大毘婆沙論》，復轉學大乘佛教，精通中觀。後因呂光、姚興之厚招，至、酒豐戒俱備。後秦弘始三年(401)抵長安，於西明園、逍遙園開始譯經，一生所譯經論七十四部。有《金剛經》、《阿彌陀經》、《法華經》、《大智度論》等。弟子有僧肇、道生、道融等。(佛光：6709；佛百：4928.2；http://zh.wikipedia.org/zh-tw/鳩摩羅什 · 2011.12.05)
 註解(完整) Notes(full) :

圖 2 「鳩摩羅什」規範資料。圖片取自「佛學規範資料庫」(<http://dev.ddbc.edu.tw/authority/person/>)。法鼓文理學院版權所有。

法鼓文理學院製作電子佛典 CBETA 及相關知識整理與開發，有十分豐富、細膩、深刻及優越的成績。最難能可貴的是，他們的資料庫、資料本身與各種成果都完全置

放網上，提供無償使用與下載。像他們的人名、地名、時間規範資料庫，即可作為自動搜尋並標示的工具，讓其他古籍資料庫的標示作業事半功倍。文本開放在此突顯為一議題。佛教團體建置資料庫，有與其他古籍製作相異之用心與目標，故不能一以概之。但誠如國學網尹小林（2013）先生所言「我們正處於一個偉大的時代，需要產生偉大的作品，創造偉大的成果」（光明日報，2013）。基於這樣的使命，就需要有宏闊的思維，有帶領古籍過渡為數位資料的責任感。

再以維基百科與臉書為例，前者為內容開放的百科全書，由全世界人類共同參與書寫，無償提供網路自由檢閱，目前中文版已有 79 萬 234 個條目（維基媒體基金會，2015 年 1 月 17 日），同時間大英百科中文版的條目為 7 萬個（遠流出版公司、智慧藏科技學習公司，無日期），相差十餘倍。維基百科不完全以多取勝，它的內容在全球協力下已愈來愈完善、深入，且免費、便利使用，這樣的成果簡直是人類的奇蹟與偉大象徵。維基尚有文庫，也是由世界各地人們自由、義務性地提供文本，文檔已有 11 萬 5,224 篇（維基媒體基金會 a，2014.10.12），不乏古籍作品。或許這些文本內容不盡精確，但它檢索方便、使用無償，查閱率不下那些收費資料庫。假以時日，這種開放式文本的成長將不可限量，有哪個團隊的組員能有幾十億人口呢？在數位時代，資料就是話語權，這是文化霸權的掌握，不啻一場戰爭。另如風靡全球的社群網站—臉書，全球活躍用戶數在 2012 年 10 月突破 10 億（維基媒體基金會 b，2014.10.12），它不收費，卻為創始人馬克·祖克柏（Mark Elliot Zuckerberg）帶來鉅額財富。祖克柏早有機會賣出此贏利商品，但他說自己關心的是「如何讓世界更開放」（維基媒體基金會 c，2014.10.12）。可見投資報酬率的資料庫使用思維有另種可能。

數位化人力與金錢資本都鉅大，不可能全部公益，但至少以國家、政府經費製作的應該向全民開放，資源共享。文本開放後提供大眾增值，增值成果若也開放，全球受益，可吸引廣泛使用；不開放的增值成果不便利利用，難以敵對免費資源。現階段古籍數位技術還相當多且專業，例如「語意檢索」（葉健欣，2009）與「視覺化古籍校勘平台」（葉健欣，2014），古籍自動標點、自動比對與自動排版技術（光明日報，2013），已遠過文史學者的可能想像。技術飛躍進步，若能再擁有文本，古籍數位化趕上二十年，或許能夠期待。

(二) 新型態的「船堅礮利」？

數位文本進行標示，後續就能展開各種運算與分析，將詞彙資料發展成視覺化訊息及衍生資料統計後的新研究議題。以目前製作相當先進與優越的法鼓文理學院佛學相關資料庫來看，他們建置了完善的知識規範資料庫並對各數位文本做了標示。標示後的資料可提供詞彙查詢，電腦自動計算詞彙出現數量與出處，還可觀看詞彙前後用字的情況與頻率（法鼓文理學院，無年代），如圖 3 及圖 4 所示。



圖 3 CBETA 檢索結果 (1)。圖片取自「佛學規範資料庫」(<http://140.112.26.229/cbetalexicon/concordance.py?term=%E7%9D%A1%E8%A6%BA>)。法鼓文理學院版權所有。

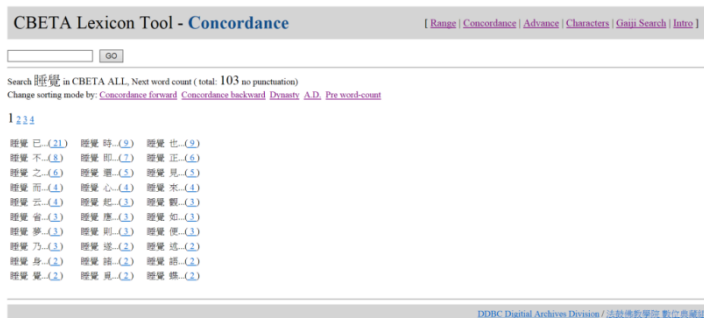


圖 4 CBETA 檢索結果 (2)。圖片取自「佛學規範資料庫」(<http://140.112.26.229/cbetalexicon/concordance.py?term=睡覺&idx=cbeta&opt=next>)。法鼓文理學院版權所有。

有了資料點，可發展 GIS 地理資訊系統，近年又開發出社會網絡分析與時間軸展現，如法鼓文理學院所製作佛教人物傳記三系統，再如臺灣清華大學中文系教授祝平次 (2014)，也利用標注古籍展開詞頻分布與視覺化的觀察與理解。當資料為平面、

敘述為文字時，我們不易察覺其間關連性與密切度，進行了可視化製作後，特別的分佈與交集便展現了。此外，校勘技術，可將兩種以上版本疊放比對，例日本所藏中文古籍數據庫「四庫提要」（據商務印書館排印本打字）與北京中華書局標點本《四庫全書總目提要》，二數位全文經疊置後，差異之處以不同顏色標出（引文底線標示），我們即刻可發現兩版本錯字及標點差異情形。如果以之利用在古籍版本間的校勘，電腦精確度強過人腦，人腦智慧大可全力在解決差異的解釋與處理，而非逐字逐卷地疲於往返文本。

校勘有新工具，輯佚也有新機會。模糊檢索技術，可跳脫全文檢索單純字詞的符合搜尋，縱橫上下句間尋找相符度。模糊檢索很早即開發，但運用在古籍的不多見。由於典籍中引用常與傳世版本內容文字參差，使用全文檢索無法找出相關資料。模糊檢索可勾稽上下幾句間集中出現的字詞。而古籍資料具有相當範圍性，類同的資料不致失控，出現過多雜訊。

那麼，所有的關鍵還是指向文本取得，以公家經費製作的數位文本應開放全民共享，讓研究者有機會利用廣大的數位資料來進行標示、製作，展開前所未有的古籍整理與研究新局。

（三）完全統計與完全分析，解開古籍密碼？

《大數據》一書中分析了過往研究法與觀念，抽樣的素材、統計的運算、假設的概念、推估的結果，但在數位資料時代，蒐集、觀察、推導的原始材料很可能是完整的、全面的，樣本就是全部母體（麥爾荀伯格、庫基耶，2013／林俊宏譯，2013，頁 40-47）。這樣龐大繁多的資料當然存在雜訊，只是當它的訊息量遠過個人能力蒐羅範圍，雜亂的部分就顯得其次，主要的指向成為難以忽略的「事實」。嚴耕望的精勤，爬梳《全唐詩》及眾多唐代史料，畫出唐代驛站設置，然在今天數位文本標示下，可能是有唐及有宋一代的全文獻都在取材範圍，唐驛站的故事或許將更加真切，尤其，今天幾乎沒有第二個嚴耕望。

數位文本經標示，可以進行文本間大串連，不受全文檢索字符之限，一個新的古籍閱讀模式也正式浮現。過去我們必須經由學習與指導，才能了解延伸閱讀的方向，知道了也必須另外勤查他書，但標示把關連性結合起來，閱讀時隨時可以連結典故出處、相關典籍乃至過去認為不一定直接有關的典籍。節省查找辛勞是小事，從這些無限可能的結合中發現問題才是重點，這可能是過去治學方法中完全想像不到的關連與

現象。以筆者嘗試建置古籍叢書目錄為例（如圖 5），所有書名、人名經標示，在智能型檢索功能輔助下，輸入一「農」字，叢書名與子目書名中有其字的全部出現，提供點選，也可以再以書名、人名詞彙為查詢值，展列相關訊息，僅有全文檢索是搜尋不全的。



圖 5 於古籍叢書目錄中以「農」為查詢詞之檢索結果。圖片取自「中國古籍叢書目錄檢索系統」(http://120.126.128.164:81/SOCRB/index2_old.aspx)。

叢書書目經標示，還可進行量化觀察。比如《中國叢書綜錄》收錄子目七萬多條，去其重複有 3 萬 8,891 種（上海圖書館編，2007）；《中國叢書廣錄》子目 5 萬 780 種，去其重複有 4 萬 227 種（陽海清，1999）。兩目合併後，不重複的子目書，經初步計算為 7 萬 1,565 本，其中僅出現過一次的有 5 萬 1,278 本。71,565 本的數量還要進一步細膩處理，去除子目書名異體字、題名小異及其他情況，估計可得 7 萬餘種，這應該可說是古典叢書收錄的子目總量。另，過去很難掌握的叢書本與單行本間重複的情況，現在首度獲得了解契機。七萬多本子目書中有五萬多本是僅見單部叢書收錄的，雖然不能遽以為傳世唯一作品，但如果連叢書中都只單見，其它刊行的可能性亦不大，至少，它整體數值之多，值得引發我們深思。

數位化給予了我們有史以來全盤掌握傳世史料的機會，完全統計、完全分析，竟真的可以做到最大化，那個「通古今之變」的期許，是有可能實現的，並非徒然的壯語與豪情。記錄並至今隱藏在書冊中的古籍密碼、古代真象、古人情懷，也許我們正重新要展開探勘與認識。

五、結論

現今《隋書·經籍志》中著錄的唐初作品，多半已不見傳世；兩唐書記載的文獻，也多有自此於歷史中消失；就是《宋史·藝文志》，還是有不少書籍我們再也無從認識。書籍散亡的原因當然很多，這個時期又不乏戰亂兵燹。但就是因為書的複本量少，一有閃失即萬劫不復。能夠流傳到後世的，有時不一定是當時最優越的作品，而是有機會被延續的作品。猶如宋樓鑰故事中的名醫陸從老，南宋像他那樣家藏珍秘醫學典籍、罕少示人的，應該不少。陸醫的秘本或許終究深藏未露，而後人見到的卻是當時坊間市面的通俗本，今天且為我們評定為古籍最高等級的「善本」。綜上所論，可分三點作結：

(一) 紙本不死，只是凋零

雕板印書終究成為書籍流通的主要方式，認為內容錯誤多也好，編刊不負責任也好，終究擋不住它迅猛發展氣勢，原因無他，符合人性而已。書籍的錯誤，解決辦法是催生校讎學。那麼，古籍、今籍全面數位化的時間也會到來，它會比雕板印書取得還要徹底，因為今人本來就沒有校讀習慣。紙本不會消失，它只是罕被「看見」。今天我們認為功力累積的紙本才是真正有價值、具深度、不可磨滅的作品，但問題是，這已非中年的我們在做判斷與選擇，而是現在二十歲的年輕人以及更年輕的下一代，他們在電子產品的世界中成長，與我們的背景迥異，兩代人已彼此無法談「價值」。世界是電腦資訊構築的。因此，紙本不見得因優良而被留下，資料庫也不因粗劣而被淘汰，真正淘汰的機制是有無數位化以及是否具相當程度的開放。目前已有不少學者認識到這個趨勢，而建置個人網站，將自己全數作品，包括學術專著與一般雜文，乃至演講與上課錄影，全數連結上網，提供無償觀看、下載。在數位時代，出名與留名的方式是免費開放，它吸引最多點閱率，點閱率及轉載愈多，「市占率」便是一種發言權。奇妙的是，一些文本明明已全部上網，實體書卻反而銷售得更好，新世界真的形成了。

(二) 頂尖的文獻人才與頂尖的技術人才

猶記 2013 年程毅中、尹小林、馬建農三位先生先後在光明日報的討論，對於古籍數位化核心應是古籍整理還是資訊技術有不同意見。個人認為，數位化要成功，古

籍要利用新技術往下傳承，古籍整理與資訊技術一樣關鍵，最優秀的古典文獻人才與最頂尖的電腦人才並重，缺一不可。古籍數位化的成果最終主要是古典研究者使用，資訊工程師「得魚忘筌」，因此，必須以符合古文獻學者要求為宗旨。然而，複雜情況的古文獻，非常不容易用講求明確精準的電腦資訊來表達，如果沒有相當優秀又具耐心的工程師，根本難以解決問題。近年多有展開雙專長培訓的想法與實際作為，此效果還要一段時間才能見到。或許新人類中會出現特別人才，但不是究竟辦法，雙領域人才可做橋樑，但突破兩種截然不同學科的結合，還是期待各方人才組成團隊，這兩者能力都不是朝夕可躋以及隨時易切換的，當標準要求得非常高時。因此，古典研究者若不滿意數位製作，應該積極參與新工作，提出問題、說明需求，甚至貢獻心力，一同加入討論、設計古籍數位化的規範與創意。這是另一個吾輩定則定矣的時代。

（三）學術 M 型化時代來臨

吳夏平（2012）曾提出「數字鴻溝」、「學術異化」，指能不能使用數位成果，將形成社會以及學術研究上的落差。祝平次（2014）教授也講「學術不平等」，指能不能擁有數位全文以及進行數位解讀，將造成學術的不平等。他們都昭告了一個學術 M 型化時代的來臨。電腦是人腦的延伸，數位工具是人類能力的放大鏡。幾十年來人類本身的力量沒有大幅改變，但工具徹底異化了人類的可能，「大數據」讓人類成為一個全知者，全世界以及各個社會的隱性密碼，過去即使借助研究方法都無法窮盡的了解與掌握，今天易如反掌，而且是坐等資料蜂擁而來。但它的門檻不小，以古籍研究來說，大型、功能資料庫都需付費，不是每一個學者都能長期支應購買，這使得有無公共資源形成研究上質與量的顯著差別。而數位解讀需要一定的資訊素養及工程師資源，又有條件與環境的限制，非一蹴可幾。目前，願開放資料庫與釋出基底資料的很少，還是如同社會救助，對貧富差距改善有限。如果持續商品化下去，那學術富者愈富、貧者愈貧的情況將很難避免。

今天，批評過度依賴數位資料導致學術研究低落的聲音不小，尤其是從實體書時代訓練過來的學者，對數位化造成的淺薄化深感無奈。但，淺閱讀是人類在面對資訊爆炸時的必然心理機轉，可以省卻勞心勞力，便利、迅速獲得成績，何樂不為？！這是人性。或者說，不一定是數位化造成的危害，是世界先淺薄化了，人淺薄化了，做研究當然也逃不掉膚淺。數位化的巨輪必然持續滾動下去，價值就待後人判斷，今天且完成一個階段性任務，為古籍傳承努力盡份心力。

誌謝

本文撰寫，電腦資訊技術與相關理念，多蒙剎那工坊葉健欣先生指導與圖版提供，在此敬表最大謝忱。

參考文獻

(傳統文獻)

- 漢·許慎(無年代)。《說文解字》(卷末宋太宗雍熙三年十一月中書門下牒徐鉉等〈新校定說文解字〉)。檢自「中國基本古籍庫」掃描《四部叢刊》景北宋本圖版。
- 唐·李延壽(1985a)。《南史》(卷41,頁1038)。臺北市:鼎文書局。
- 唐·李延壽(1985b)。《南史》(卷59,頁1460)。臺北市:鼎文書局。
- 唐·房玄齡等(1983a)。《晉書》(卷72,頁1911)。臺北市:鼎文書局。
- 唐·房玄齡等(1983b)。《晉書》(卷75,頁1982)。臺北市:鼎文書局。
- 唐·姚思廉等(1983)。《梁書》(卷33,頁486)。臺北市:鼎文書局。
- 唐·封演(無年代)。《封氏聞見記》(卷2〈石經〉條,葉七)。檢自「中國基本古籍庫」掃描明雪晴齋抄本圖版。
- 後晉·劉煦等(1985a)。《舊唐書》(卷165,頁4307)。臺北市:鼎文書局。
- 後晉·劉煦等(1985b)。《舊唐書》(卷84,頁2802)。臺北市:鼎文書局。
- 宋·孔延之(無年代)。《會稽掇英總集》(卷1〈會稽掇英總集原序〉)。檢自「中國基本古籍庫」掃描清文淵閣四庫全書補配清文津閣四庫全書本圖版。
- 宋·王溥(1998)。《五代會要》(卷8〈經籍〉條,頁96)。北京市:中華書局。
- 宋·王應麟(無年代a)。《玉海·藝文》(卷55〈祥符賜莊列子〉)。檢自「中國基本古籍庫」掃描元至元慶元路儒學刻明遞修本圖版。
- 宋·王應麟(無年代b)。《玉海·食貨》(卷178〈景德農田敕〉)。檢自「中國基本古籍庫」掃描元至元慶元路儒學刻明遞修本圖版。
- 宋·王應麟(無年代c)。《玉海·藝文》(卷63〈天聖校定內經素問〉)。檢自「中國基本古籍庫」掃描元至元慶元路儒學刻明遞修本圖版。
- 宋·王應麟(無年代d)。《玉海·帝學》(卷27〈景德國子監觀羣書漆板〉)。檢自「中國基本古籍庫」掃描元至元慶元路儒學刻明遞修本圖版。
- 宋·司馬光(無年代)。《溫國文正司馬公文集》(卷62〈與劉道原書〉)。檢自「中國基本古籍庫」掃描《四部叢刊》景宋紹興本圖版。
- 宋·朱彧(無年代)。《萍洲可談》。檢自「中國基本古籍庫」掃描《百川學海》本圖版。

- 宋·朱翌（無年代）。*猗覺寮雜記*（卷上）。檢自「中國基本古籍庫」掃描清知不足齋叢書本圖版。
- 宋·李昉（無年代）。*文苑英華*（卷 10 盧肇〈天河賦〉末識）。檢自「中國基本古籍庫」掃描明刻本圖版。
- 宋·李觀（2004）。*直講李先生文集*（卷 25〈《皇祐續藁》序〉，頁 176）。北京市：綫裝書局。
- 宋·李燾（1961a）。*續資治通鑒長編*（卷 60〈真宗〉，葉一）。
- 宋·李燾（1961b）。*續資治通鑒長編*（卷 74〈真宗〉，葉十三）。
- 宋·李燾（1961c）。*續資治通鑒長編*（卷 61，葉十）。
- 宋·李燾（1961d）。*續資治通鑒長編*（卷 61，葉九）。
- 宋·李燾（1961e）。*續資治通鑒長編*（卷 445，葉十一）。
- 宋·周必大（2004）。*廬陵周益國文忠公集*（卷 55，頁 555）。北京市：綫裝書局。
- 宋·周輝（1981）。*清波雜誌*（卷 4，頁 21）。臺北市：臺灣商務印書館。
- 宋·姚鉉編（無年代）。*唐文粹*（卷末附〈唐文粹後序〉）。檢自「中國基本古籍庫」掃描《四部叢刊》景元翻宋小字本圖版。
- 宋·徐鉉（無年代）。*徐公文集*（卷 23〈韻譜後序〉）。檢自「中國基本古籍庫」掃描《四部叢刊》景黃丕烈校宋本圖版。
- 宋·張詠（無年代 a）。*乖崖先生文集*（卷 8〈許昌詩集序〉）。檢自「中國基本古籍庫」掃描《續古逸叢書》本景宋本圖版。
- 宋·張詠（無年代 b）。*徐公文集*（卷末附晏殊〈後序〉）。檢自「中國基本古籍庫」掃描《續古逸叢書》本景宋本圖版。
- 宋·陸游（無年代 a）。*渭南文集*（卷 30〈跋六一居士《集古錄》跋尾〉）。檢自「中國基本古籍庫」掃描宋嘉定刻本圖版。
- 宋·陸游（無年代 b）。*渭南文集*（卷 28〈跋唐盧肇集〉）。
- 宋·陸游（無年代 c）。*渭南文集*（卷 26〈跋歷代陵名〉）。
- 宋·程俱（無年代 a）。*麟臺故事*（卷 2 中〈校讎〉）。檢自「中國基本古籍庫」掃描《四部叢刊續編》景明景宋鈔本圖版。
- 宋·程俱（無年代 b）。*麟臺故事*（卷 2 中〈書籍〉）。檢自「中國基本古籍庫」掃描《四部叢刊續編》景明景宋鈔本圖版。
- 宋·葉夢得（無日期）。*石林燕語*（卷 8）。檢自「中國基本古籍庫」掃描明正德楊武刻本圖版。
- 宋·劉跂（無日期）。*學易集*（卷 6）。檢自「中國基本古籍庫」掃描清《武英殿聚珍版叢書》本。

- 宋・樓鑰 (1979a)。*攻媿集* (卷 71〈跋華氏《中藏經》〉)。臺北市：臺灣商務印書館。
- 宋・樓鑰 (1979b)。*攻媿集* (卷 77〈跋春秋繁露〉)。臺北市：臺灣商務印書館。
- 宋・歐陽修 (無年代)。*歐陽文忠公文集* (卷 73〈外集〉卷 23〈跋永成縣學記〉，葉十四)。檢自「中國基本古籍庫」掃描宋慶元二年周必大刻本圖版。
- 宋・歐陽修等 (1985)。*新唐書* (卷 163，頁 5025)。臺北市：鼎文書局。
- 宋・蔡襄 (無日期)。*莆陽居士蔡公文集* (卷 25〈《茶錄》後序〉)。檢自「中國基本古籍庫」掃描宋刻本圖版。
- 宋・蘇軾 (無日期 a)。*東坡先生志林* (卷 5)。檢自「中國基本古籍庫」掃描明刻本。
- 宋・蘇軾 (無日期 b)。*蘇文忠公全集* (卷 32〈李氏山房藏書記一首〉)。檢自「中國基本古籍庫」掃描明成化本圖版。
- 元・脫脫等 (1983)。*宋史·藝文志* (頁 5033)。

(近人論著)

- 上海圖書館編 (2007)。*中國叢書綜錄·子目*。上海市：上海古籍出版社。
- 中央研究院歷史語言研究所 (無年代)。*漢籍電子文獻資料庫*〔資料庫〕。檢自 <http://hanchi.ihp.sinica.edu.tw/ihpc/hanjiquery?19^954109311^10^^^@@@558717561>
- 孔凡禮點校 (2005)。*唐書判*。在宋・洪邁，*容齋隨筆* (卷 10，頁 129)。
- 尹小林 (2013.5.28)。*古籍數位化應以技術為突破口—兼與程毅中先生商榷*。*光明日報*。檢自 <http://www.guoxue.com/?p=12237>
- 毛遠明 (2009)。*碑刻文獻學通論* (第三章，頁 279-289)。北京市：中華書局。
- 王瑞來點校 (1983)。*手寫九經*。在宋・羅大經撰，*鶴林玉露* (卷 1，頁 3)。
- 北京國學時代文化傳播有限公司 (無年代)。*國學寶典*〔資料庫〕。檢自 <http://www.gxbd.com/>
- 光明日報 (2013.9.16)。*國學大數據時代來了*。*光明日報*，15 版。檢自 <http://www.guoxue.com/?p=14719>
- 佛學規範資料庫〔資料庫〕。新北市：法鼓文理學院。
- 邵正坤 (2014)。*古籍數字化的困局及應對策略*。*圖書館學研究*，12，32-34，79。
- 吳夏平 (2009.8.18)。*論古籍數位化與現代學術進程*。在中國首都師範大學電子文獻研究所與首都師範大學國學傳播中心共同主辦，*第二屆中國古籍數位化國際學術研討會*，檢自 <http://www.guoxue.com/wk/000662.htm>
- 吳夏平 (2012)。*古籍數字化與學術異化*。*山西師大學報*，39(5)，131-135。
- 李更 (2006)。*宋代館閣校勘研究*。南京市：鳳凰出版社。
- 李瑞良 (2000)。*中國古代圖書流通史*。上海市：上海人民出版社。

- 來新夏等（1990）。*中國古代圖書事業史*。上海市：上海人民出版社。
- 周寶榮（2003）。*宋代出版史研究*（第六章，頁 133-176）。鄭州市：中州古籍出版社。
- 法鼓文理學院（無年代）。*CBETA Lexicon Tool*。檢自 <http://140.112.26.229/cbetalexicon/concordance.py?term=%E7%9D%A1%E8%A6%BA>
- 范佳（2013）。「數字人文」內涵與古籍數字化的深度開發。*圖書館學研究*，3，29-32。
- 祝平次（2014.6.5）。文獻標注與文史研究。在中央研究院數位文化中心（主辦），*古籍全文資料庫的回顧與展望工作坊*。中央研究院數位文化中心，臺北市。取自 <http://xiaoxue.iis.sinica.edu.tw/workshop/PDF/04-%E6%96%87%E7%8D%BB%E6%A8%99%E6%B3%A8%E8%88%87%E6%96%87%E5%8F%B2%E7%A0%94%E7%A9%B6.pdf>
- 馬建農（2013.6.25）。古籍數位化的核心是文化內涵。*光明日報*。取自 http://epaper.gmw.cn/gmrb/html/2013-06/25/nw.D110000gmrb_20130625_2-13.htm
- 馬創新、曲維光、陳小荷（2014）。中文古籍數字化的開發層次和發展趨勢。*圖書館*，2，104-106。
- 張秀民著，韓琦增訂（2006）。*中國印刷史*。杭州市：浙江古籍出版社。
- 張茂鵬點校（1983）。書籍之厄（頁 217）。在宋·周密，*齊東野語*（卷 12）。
- 張靜、王兆鵬（2010）。北宋文人對新舊傳播方式的態度與選擇—兼與當代數字化出版浪潮比較。*中州學刊*，6，199-203。
- 曹之（1994）。*中國印刷術的起源*。武漢市：武漢大學出版社。
- 曹之（2007）。*中國古籍版本學*。武漢市：武漢大學出版社。
- 陳德弟（2004）。備書業的興衰和雕版印刷術的發明。*出版科學*，5，67-70。
- 陸宇傑、許鑫、郭金龍（2012）。文本挖掘在人文社會科學研究中的典型應用述評。*圖書情報工作*，56(8)，18-25。
- 麥爾荀伯格（Mayer-Schonberger, V.）、庫基耶（Cukier, K.）著（2013）。*大數據*（林俊宏譯）。臺北市：天下文化。（原作 2013 年出版）
- 曾棗莊、劉琳主編（2006）。*全宋文*（卷 78，頁 407）。上海市：辭書出版社。
- 程毅中（2013.4.30）。古籍數位化須以古籍整理為基礎。*光明日報*。取自 <http://www.guoxue.com/?p=12237>
- 陽海清（1999）。*中國叢書廣錄*。武漢市：湖北人民出版社。
- 黃河（2014）。古籍數字化利弊芻議。*新世紀圖書館*，5，27-29。
- 葉健欣（2009.5）。*剎那人文系統：一個以語意為核心的文本處理平台*。檢自 <https://docs.google.com/document/pub?id=1kVOejyoUshM46ih4itI5FpKPmOV B5WVS9E9tYDvLBpY>

葉健欣 (2014.6.5)。視覺化古籍校勘平台—人文與科技協作之嘗試。在中央研究院數位文化中心 (主辦)，*古籍全文資料庫的回顧與展望工作坊*。中央研究院數位文化中心，臺北市。檢自 <http://xiaoxue.iis.sinica.edu.tw/workshop/PDF/05-%E8%A6%96%E8%A6%BA%E5%8C%96%E5%8F%A4%E7%B1%8D%E6%A0%A1%E5%8B%98%E5%B9%B3%E5%8F%B0.pdf>

維基媒體基金會 (2014.10.12a)。維基文庫。檢自 <http://zh.wikisource.org/zh-hant/Wikisource:%E9%A6%96%E9%A1%B5>

維基媒體基金會 (2014.10.12b)。Facebook。檢自 <http://zh.wikipedia.org/zh-tw/Facebook>

維基媒體基金會 (2014.10.12c)。馬克·祖克柏。檢自 <http://zh.wikipedia.org/wiki/%E9%A9%AC%E5%85%8B%C2%B7%E6%89%8E%E5%85%8B%E4%BC%AF%E6%A0%BC>

維基媒體基金會 (2015.1.17)。維基百科：歡迎。檢自 <http://zh.wikipedia.org/wiki/Wikipedia:%E6%AC%A2%E8%BF%8E>

遠流出版公司、智慧藏科技學習公司 (無日期)。大英線上/*Britannica Online*。檢自 http://www.ylib.com/britannica/online_chinese8/

謝清俊 (2007)。後設資料與內容標誌 (上)：淺談數位化的文章和其內容標誌處理。*佛教圖書館館刊*，45。檢自 <http://www.gaya.org.tw/journal/m45/45-lib1.pdf>

謝清俊 (2008)。後設資料與內容標誌 (下)：意義的處理，以心經為例。*佛教圖書館館刊*，47。檢自 <http://www.gaya.org.tw/journal/m47/47-lib3.pdf>

謝清俊 (2014.9.2)。淺談人文資訊學的回顧與省思。在中國科學院，第10屆兩岸三院資訊技術與應用交流研討會。中國科學院，北京市。檢自 <http://2a3y2014.csp.escience.cn/dct/page/70002>

嚴耕望 (1956)。唐僕尚丞郎表。臺北市：中央研究院歷史語言研究所。

嚴耕望 (2008)。治史三書 (頁 135)。上海市：上海人民出版社。

顧青 (2013.8.29)。中文古籍數位化的發力點。*中國新聞出版報*。檢自 <http://data.chinaxwcb.com/epaper2013/epaper/d5608/d8b/201308/36696.html>